





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:57:55 ; Search time 198.55 Seconds

(without alignments)  
7.834 Million cell updates/sec

Title: US-09-290-049A-11

Perfect score: 116  
Sequence: 1 AINHLSTLEAMSNDPOYNKD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

1 number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq-1101.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	90	77.6	1577	17	AA91047
2	74	63.8	1592	14	AA92925
3	64	55.2	2057	21	AA10667
4	50	43.1	473	21	AA92078
5	50	43.1	474	21	AA92077
6	50	43.1	522	21	AA92076
7	49	42.2	418	22	AA93916
8	49	42.2	531	22	AA92987
9	49	42.2	623	20	AA98707
10	49	42.2	760	22	AA96537
11	47	40.5	195	21	AA90941

#### ALIGNMENTS

12	46	39.7	371	21	AA931391	Arabidopsis thalia
13	46	39.7	372	21	AA931390	Arabidopsis thalia
14	46	39.7	411	21	AA931378	Arabidopsis thalia
15	46	39.7	412	21	AA931377	Arabidopsis thalia
16	46	39.7	420	21	AA931389	Arabidopsis thalia
17	46	39.7	460	21	AA931376	Arabidopsis thalia
18	46	39.7	826	22	AA981986	S. epidermidis ope
19	45.5	39.2	560	21	AA938644	Arabidopsis thalia
20	45	38.8	151	20	AA935397	Chlamydia pneumonia
21	45	38.8	409	21	AA944663	Zea mays protein f
22	45	38.8	410	21	AA943434	Zea mays protein f
23	45	38.8	410	21	AA944662	Zea mays protein f
24	45	38.8	458	21	AA934333	Zea mays protein f
25	45	38.8	513	21	AA934332	Zea mays protein f
26	45	38.8	513	21	AA934332	Zea mays protein f
27	45	38.8	657	14	AA932657	PSI protein from C
28	45	38.8	657	22	AA929116	C glutathione prote
29	44.5	38.4	259	17	AA989326	Periplasmic chap
30	43.5	37.5	197	19	AA950026	Human interleukin-
31	43.5	37.5	197	21	AA957326	Mature human IL-4R
32	43.5	37.5	498	22	AA943563	Human polypeptide
33	43.5	37.5	691	21	AA92202	Fusion polypeptide
34	43.5	37.5	694	21	AA92201	Fusion polypeptide
35	43.5	37.5	694	21	AA92203	Fusion polypeptide
36	43.5	37.5	735	22	AA900394	Human secreted pro
37	43.5	37.5	739	22	AA962397	Human M8SP1 polype
38	43.5	37.5	739	22	AA962398	Human M8SP2 polype
39	43.5	37.5	784	21	AA92207	IL-13/IL-4 dual tr
40	43.5	37.5	793	21	AA92208	Derived amino acid
41	43.5	37.5	824	11	AA904575	Human IL-4 recepto
42	43.5	37.5	825	12	AA911254	Human interleukin-
43	43.5	37.5	825	18	AA913499	Human interleukin-
44	43.5	37.5	825	19	AA960668	Human interleukin-
45	43.5	37.5	825	19	AA948314	Human interleukin-

#### RESULT 1

ID	AA91047	standard; Protein; 1577 AA.
AC	AA91047;	
DT	22-MAY-1996	(first entry)
DE	Alpha-D-glucosyltransferase.	
KW	Alpha-D-glucosyltransferase; primer-independent; soluble glucan;	
KW	sucrose; transgenic plant; cloning; Escherichia coli;	
KW	phage lambda-C13; vector; plasmid pSG501; plasmid pSG502;	
KW	gene transfer; crop improvement; storage carbohydrate; pasture;	
KW	feedstuff; senescence; dextran; binder; food; pharmaceutical.	
OS	Streptococcus salivarius strain ATCC 25975.	
XX		
XX	WO9606173-A1.	
XX		
XX	29-FEB-1996.	
XX		
XX	24-AUG-1995;	95WO-AU00527.
XX		
XX	24-AUG-1994;	94AU-0007643.
XX		
XX	(GTFP) GTFPARD P M.	
XX	(JACO) JACOUES N A.	
XX	(STMP) STIMPSON C L.	
XX		
XX	Giffard PM, Jacques NA, Simpson CL;	
XX	WPI: 1996-151376/15.	
XX		
XX	N-PSDB: AAT13139.	





OY 1 AINHSLIAMSNDPOY 18  
| ||||| | : ||||  
Db 665 ankhlstledwngkdpqy 682

## RESULT 4

AAAG42078  
ID AAAG42078 standard; Protein: 473 AA.

AC AAAG42078;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 52432.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.

XX Arabidopsis thaliana.

PR EP1033405-A2.

PD 06-SEP-2000.

PE 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 18-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136382.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 21-JUN-1999; 99US-0139763.

PR 22-JUN-1999; 99US-0139817.

PR 23-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 03-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0148426.  
PR 20-AUG-1999; 99US-0148722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155148.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0158294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161933.  
PR 29-OCT-1999; 99US-0162142.

Query Match Best Local Similarity 43.1%; Score 50; DB 21; Length 473;

Matches 11; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

QY 1 AINHSLEAWSDN--DPOYKND 21  
DB 160 algnavllstswenygdpndend 182

RESULT 5

AAG42077  
ID AAG42077 standard; Protein; 474 AA.  
XX  
AC AAG42077;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 52431.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PE 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140931.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 09-AUG-1999; 99US-0148171.  
PR 10-AUG-1999; 99US-0148319.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.

PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154038.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 43.1%; Score 50; DB 21; Length 474;  
Best Local Similarity 47.8%; Pred. No. 10;  
Matches 11; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

Oy 1 AINHSILEAFNSDN--DPOYNKD 21  
|||:|||||  
Db 161 algnavlsswsgnypgdpnskd 183

RESULT 6  
AA642076 ID AA642076 standard; Protein: 522 AA.

XX AC AA642076;

XX DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 52430.  
XX  
KW Protein identification: signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123160.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147304.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.

PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154039.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155149.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 43.1%; Score 50; DB 21; Length 522;  
Best Local Similarity 47.8%; Pred. No. 12;  
Matches 11; Conservative 3; Mismatches 7; Indels 2; Gaps 1;  
QY 1 AINHLSTLEAWSND--DPQYKND 21  
|||:|||||||  
Db 209 aigmavilsswsenygdpksenkd 231

RESULT 7  
ID AAB93916 standard; Protein: 418 AA.  
AC AAB93916;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
De Human protein sequence SEQ ID NO:13888.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
OS Homo sapiens.  
XX  
PN EP1074617-A2.

XX 07-FEB-2001.  
PD  
XX 28-JUL-2000; 2000EP-0116126.  
PF  
XX 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI: 2001-318749/34.  
DR  
XX  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX  
PS Claim 8: SEQ ID 13888; 2537bp + CD ROM; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 418 AA;

Query Match 42.2%; Score 49; DB 22; Length 418;  
Best Local Similarity 52.9%; Pred. No. 13;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
QY 5 ISILEAWSNDPQYKND 21  
::||:|||||||  
Db 107 imvlagasefdpqnkd 123

RESULT 8  
ID AAB92987 standard; Protein: 531 AA.  
AC AAB92987;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
De Human protein sequence SEQ ID NO:11710.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
OS Homo sapiens.

```

PM      EPI074617-A2.
XX
XX      07-FEB-2001.
XX
XX      28-JUL-2000; 2000EP-0116126.
XX
XX      29-JUL-1999; 99JP-0248036.
PR      27-AUG-1999; 99JP-0300253.
PR      11-JAN-2000; 2000JP-0118776.
PR      02-MAY-2000; 2000JP-0183767.
PR      09-JUN-2000; 2000JP-0241899.
XX
XX      (HELI-) HELIX RES INST.
PA
XX      Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI      Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX      WPI; 2001-318749/34.
XX
XX      Primer sets for synthesizing polynucleotides, particularly the 5602
PT      full-length cDNAs defined in the specification, and for the detection
PT      and/or diagnosis of the abnormality of the proteins encoded by the
PT      full-length cDNAs -
XX
PS      Claim 8; SEQ ID 117A10; 2537P + CD ROW; English.
XX
XX      The present invention describes primer sets for synthesizing 5602
CC      full-length cDNAs defined in the specification. Where a primer set
CC      comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC      to the complementary strand of a polynucleotide which comprises one of
CC      the 5602 nucleotide sequences defined in the specification, where the
CC      oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC      of an oligonucleotide comprising a sequence complementary to the
CC      complementary strand of a polynucleotide which comprises a 5'-end
CC      sequence and an oligonucleotide comprising a sequence complementary to a
CC      polynucleotide which comprises a 3'-end sequence, where the
CC      oligonucleotide comprises at least 15 nucleotides and the combination of
CC      the 5'-end sequence/3'-end sequence is selected from those defined in
CC      the specification. The primer sets can be used in antisense therapy and
CC      in gene therapy. The primers are useful for synthesizing polynucleotides,
CC      particularly full-length cDNAs. The primers are also useful for the
CC      detection and/or diagnosis of the abnormality of the proteins encoded by
CC      the full-length cDNAs. The primers allow obtaining of the full-length
CC      cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
CC      AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC      AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
        represent oligonucleotides, all of which are used in the exemplification
        of the present invention.
XX
SQ      Sequence      531 AA:
XX
        Query Match          42.2%; Score 49; DB 22; Length 531;
        Best Local Similarity 52.9%; Pred. No. 17;
        Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0
OY      5 LSTLEAKSNDPOYNKD 21
       : : | : |||||
Db      247 lmvlagasefdpqykd 263

RESULT_ 9
ID      AAW88707 standard; Protein: 623 AA.
XX
XX      AAW88707;
AC
XX
XX      01-MAR-1999 (first entry)
DT
XX
DE      Secreted protein encoded by gene 174 clone HE9PB42.
XX
XX      Human; secreted protein; fusion protein; gene therapy; protein therapy;
KM

```

KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia
KW	developmental abnormality; foetal deficiency; blood; allergy; renal;
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW	inflammatory; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW	conjunctive disorder; schiophrenia; prostate; obesity; osteoclast; thymus;
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX	
XX	Homo sapiens.
OS	
XX	
PN	MO9854963-A2.
PD	10-DEC-1998.
XX	
PE	04-JUN-1998; 98MO-U0511422.
XX	
PR	18-DEC-1997; 97US-0070923.
PR	06-JUN-1997; 97US-0048877.
PR	06-JUN-1997; 97US-0048881.
PR	06-JUN-1997; 97US-0048884.
PR	06-JUN-1997; 97US-0048893.
PR	06-JUN-1997; 97US-0048896.
PR	06-JUN-1997; 97US-0048899.
PR	06-JUN-1997; 97US-0048915.
PR	06-JUN-1997; 97US-0048949.
PR	06-JUN-1997; 97US-0048964.
PR	06-JUN-1997; 97US-0048972.
PR	06-JUN-1997; 97US-0049020.
PR	06-JUN-1997; 97US-0049375.
PR	05-SEP-1997; 97US-0057628.
PR	05-SEP-1997; 97US-0057635.
PR	05-SEP-1997; 97US-0057644.
PR	05-SEP-1997; 97US-0057647.
PR	05-SEP-1997; 97US-0057650.
PR	05-SEP-1997; 97US-0057661.
PR	05-SEP-1997; 97US-0057667.
PR	05-SEP-1997; 97US-0057761.
PR	05-SEP-1997; 97US-0057764.
PR	05-SEP-1997; 97US-0057770.
PR	05-SEP-1997; 97US-0057775.
PR	05-SEP-1997; 97US-0057778.
PR	06-JUN-1997; 97US-0048875.
PR	06-JUN-1997; 97US-0048878.
PR	06-JUN-1997; 97US-0048882.
PR	06-JUN-1997; 97US-0048885.
PR	06-JUN-1997; 97US-0048894.
PR	06-JUN-1997; 97US-0048897.
PR	06-JUN-1997; 97US-0048900.
PR	06-JUN-1997; 97US-0048916.
PR	06-JUN-1997; 97US-0048962.
PR	06-JUN-1997; 97US-0048970.
PR	06-JUN-1997; 97US-0048974.
PR	06-JUN-1997; 97US-0049373.
PR	05-SEP-1997; 97US-0057584.
PR	05-SEP-1997; 97US-0057629.
PR	05-SEP-1997; 97US-0057642.
PR	05-SEP-1997; 97US-0057645.
PR	05-SEP-1997; 97US-0057771.
PR	05-SEP-1997; 97US-0057776.
PR	06-JUN-1997; 97US-0048876.
PR	06-JUN-1997; 97US-0048880.
PR	06-JUN-1997; 97US-0048883.
PR	06-JUN-1997; 97US-0048892.
PR	06-JUN-1997; 97US-0048895.
PR	06-JUN-1997; 97US-0048898.
PR	06-JUN-1997; 97US-0048901.
PR	06-JUN-1997; 97US-0048917.

PR 06-JUN-1997; 97US-0048963.  
 PR 06-JUN-1997; 97US-0048971.  
 PR 06-JUN-1997; 97US-0049019.  
 PR 06-JUN-1997; 97US-0049374.  
 PR 05-SEP-1997; 97US-0057627.  
 PR 05-SEP-1997; 97US-0057634.  
 PR 05-SEP-1997; 97US-0057643.  
 PR 05-SEP-1997; 97US-0057649.  
 PR 05-SEP-1997; 97US-0057654.  
 PR 05-SEP-1997; 97US-0057666.  
 PR 05-SEP-1997; 97US-0057760.  
 PR 05-SEP-1997; 97US-0057763.  
 PR 05-SEP-1997; 97US-0057769.  
 PR 05-SEP-1997; 97US-0057774.  
 PR 05-SEP-1997; 97US-0057777.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA;  
 Fan P, Feng P, Ferrie AM, Fischer CL, Florence C;  
 Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW;  
 Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM;  
 Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;  
 WPI: 1999-059865/05.  
 DR N-PSDB: AAW84584.  
 XX  
 PT New isolated human genes and the secreted polypeptides they encode -  
 PT useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders  
 XX  
 PS Claim 11; Page 569-571; 772pp; English.  
 XX  
 CC The invention relates to nucleic acid sequences (AAW84411 to AAW84633)  
 CC encoding human secreted proteins (AAW88534 to AAW88756). The secreted  
 CC protein gene sequences are deposited with the ATCC under deposit numbers  
 CC ATCC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,  
 CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host  
 CC cells comprising recombinant vectors containing the nucleic acid  
 CC sequences are used for the recombinant production of the secreted  
 CC proteins. The polynucleotide and amino acid sequences are useful for are  
 CC useful for preventing, treating or ameliorating medical conditions e.g.  
 CC by protein or gene therapy. Pathological conditions can be also  
 CC diagnosed by determining the amount of the new polypeptides in a sample  
 CC or by determining the presence of mutations in the new polynucleotides.  
 CC Specific uses are described for each of the polynucleotides, based on  
 CC which tissues they are most highly expressed in, and include developing  
 CC products for the diagnosis or treatment of cancer, neurodegenerative  
 CC disorders, developmental abnormalities and foetal deficiencies, blood  
 CC disorders, tumours, leukemias, diseases of the immune system, autoimmune  
 CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,  
 CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,  
 CC reitenois, prostate diseases, obesity, disorders involving osteoclasts  
 CC such as osteoporosis, arthritis or malignancies, diseases of testes, the  
 CC lung or thymus, digestive/endocrine disorders, infections and AIDS. The  
 CC polypeptides are also useful for identifying their binding partners.  
 CC The present sequence represents human secreted protein (see descriptor  
 CC line for gene number and clone identification).  
 CC  
 XX  
 SQ Sequence 623 AA;

Query Match 42.2%; Score 49; DB 20; Length 623;  
 Best Local Similarity 52.9%; Pred. No. 21;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 OY 5 LSI:AWSDNDPOYNKD 21  
 Db 107 imvixgasefdpynkd 123

RESULT 10

AA66537  
 ID AAG6537 standard; Protein: 760 AA.  
 XX  
 AC AAG6537;  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX  
 DE Human interferon-alpha induced polypeptide, SEC 63.  
 XX  
 KW Human; interferon-alpha induced gene; type I interferon treatment;  
 KW chronic viral hepatitis; relapsing remitting multiple sclerosis;  
 KW neoplastic disease; SEC 63.  
 XX  
 OS Homo sapiens.  
 XX  
 FM WO200159155-A2.  
 PD 16-AUG-2001.  
 XX  
 PF 09-FEB-2001; 2001WO-GB00578.  
 XX  
 PR 11-FEB-2000; 2000GB-0003203.  
 PR 11-FEB-2000; 2000GB-0003204.  
 PR 11-FEB-2000; 2000GB-0003205.  
 PR 11-FEB-2000; 2000GB-0003206.  
 PR 11-FEB-2000; 2000GB-0003207.  
 PR 11-FEB-2000; 2000GB-0003208.  
 PR 11-FEB-2000; 2000GB-0003210.  
 PR 11-FEB-2000; 2000GB-0003212.  
 PR 11-FEB-2000; 2000GB-0003213.  
 PR 11-FEB-2000; 2000GB-0003215.  
 PR 11-FEB-2000; 2000GB-0003220.  
 PR 11-FEB-2000; 2000GB-0003221.  
 PR 11-FEB-2000; 2000GB-0003222.  
 PR 17-FEB-2000; 2000GB-0003768.  
 XX  
 PA (PHAR-) PHARMA PACIFIC PTY LTD.  
 PI Meritet J, Dron M, Tovey MG;  
 XX  
 DR WPI: 2001-483570/52.  
 DR N-PSDB: AAW76463.  
 XX  
 PT Predicting responsiveness of a patient to treatment with a type I  
 PT interferon comprising determining the level of induced proteins after  
 PT treatment with a type I interferon, -  
 XX  
 PS Claim 1; Page 77-80; 133pp; English.  
 XX  
 CC The invention relates to a method for predicting responsiveness of a  
 CC patient to treatment with a type I interferon. The method comprises  
 CC determining the level of one or more proteins with a 646, 164, 126, 598,  
 CC 98, 177, 761, 361, 941, 657, 817, 429, 473, 399, 285 or 303 amino acid  
 CC sequence fully defined in the specification after treatment with a  
 CC type I interferon. The method allows a physician to determine whether  
 CC a patient suffering from chronic viral hepatitis, neoplastic disease  
 CC or relapsing remitting multiple sclerosis will respond favourably to  
 CC type I interferon treatment via oromucosal administration. The  
 CC present sequence is one of the polypeptides listed above that may  
 CC be used in the method.  
 CC  
 XX  
 SQ Sequence 760 AA;

Query Match 42.2%; Score 49; DB 22; Length 760;  
 Best Local Similarity 52.9%; Pred. No. 26;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 OY 5 LSI:AWSDNDPOYNKD 21  
 Db 247 imvixgasefdpynkd 263

```
RESULT 11
AAG09941
ID AAG09941 standard; Protein; 195 AA.
XX
AC AAG09941;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 8067.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI03405-A2.
XX
06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132653.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0145386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
```





```

PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141847.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.

```

```

PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 39.7%; Score 46; DB 21; Length 371;
Best Local Similarity 52.9%; Pred. No. 33;
Matches 9; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

Oy 7 ILEAWSDN--DPOYKMD 21
  ||:|:| || ||
Db 166 ILSWTENYGPSPENKD 182

RESULT 13
AAG31390
ID AAG31390 standard; Protein; 372 AA.
XX
AC AAG31390;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37688.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX

```

OS Arabidopsis thaliana.  
XX EPI033405-AZ.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137502.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149425.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.



PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147203.  
PR 04-AUG-1999; 99US-0147302.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154039.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 39.7%; Score 46; DB 21; Length 411;  
Best Local Similarity 52.9%; Pred No. 37;  
Matches 9; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 7 ILEANSN-DPOYKMD 21  
||:|:| || |||  
Db 166 llswtlenygdpsehkD 182

#### RESULT 15

ID AAG13577 standard; Protein; 412 AA.

XX AAG13577;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SPQ ID NO: 13123.

KW Protein identification: signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

OS Arabidopsis thaliana.

XX EPI033405-A2.

XX EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 28-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142802.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145066.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147492.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153738.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.

PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159328.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159337.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Query Match

Best Local Similarity 39.7%; Score 46; DB 21; Length 412;  
Matches 9; Conservativity 52.9%; Pred. No. 38;

Mismatches 3; Indels 2; Gaps 1;

QY 7 ILEAMSDN--DPQYKND 21  
||:|:| || |||  
Db 167 IISWLENYGDPSNKD 183

Search completed: March 27, 2002, 13:57:55  
Job time: 523 sec





Wed Mar 27 15:11:30 2002

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:59:29 ; Search time 87.3 Seconds

(without alignments)  
5.413 Million cell updates/sec

Title: US-09-290-049a-11

Perfect score: 116  
Sequence: 1 AINHLSTLEAWSNDNDPOYNKD 21Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

212252

Maximum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCRNUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	116	100.0	1430	3 US-09-008-172-2	Sequence 2, Appl
2	116	100.0	1430	4 US-09-210-361-6	Sequence 6, Appl
3	90	77.6	1577	2 US-08-793-824-2	Sequence 2, Appl
4	75	64.7	1475	3 US-09-007-999-2	Sequence 2, Appl
5	75	64.7	1475	4 US-09-210-361-2	Sequence 2, Appl
6	66	56.9	1375	4 US-09-210-361-4	Sequence 4, Appl
7	46	39.7	703	4 US-09-422-869-27	Sequence 27, Appl
8	45	38.8	324	3 US-08-508-761B-31	Sequence 31, Appl
9	45	38.8	657	3 US-08-508-761B-2	Sequence 2, Appl
10	44.5	38.4	259	1 US-08-277-311A-3	Sequence 3, Appl
11	44.5	38.4	259	2 US-08-473-750-6	Sequence 6, Appl
12	44.5	38.4	259	2 US-08-477-326-6	Sequence 6, Appl
13	43.5	37.5	197	3 US-08-897-020-7	Sequence 6, Appl
14	42	36.2	205	2 US-08-684-024-6	Sequence 6, Appl
15	42	36.2	205	3 US-09-145-668-6	Sequence 6, Appl
16	42	36.2	1001	1 US-07-997-556-6	Sequence 6, Appl
17	42	36.2	1001	1 US-07-943-843-2	Sequence 2, Appl
18	42	36.2	1001	1 US-08-347-003-2	Sequence 2, Appl
19	42	36.2	1097	1 US-07-943-843-6	Sequence 6, Appl
20	42	36.2	1097	1 US-08-347-003-6	Sequence 6, Appl
21	41.5	35.8	720	4 US-09-296-284-25	Sequence 2, Appl
22	41.5	35.8	754	4 US-09-296-284-4	Sequence 4, Appl
23	41	35.3	703	3 US-08-835-099A-1	Sequence 1, Appl
24	41	35.3	703	3 US-08-157-349-1	Sequence 1, Appl
25	41	35.3	712	2 US-08-835-099A-2	Sequence 2, Appl
26	41	35.3	712	3 US-09-157-349-2	Sequence 2, Appl
27	40.5	34.9	878	2 US-08-708-541A-26	Sequence 26, Appl

28	40.5	34.9	2186	2 US-08-822-445-2	Sequence 2, Appl
29	40.5	34.9	3672	2 US-08-822-445-12	Sequence 12, Appl
30	40.5	34.9	3801	2 US-08-822-445-10	Sequence 10, Appl
31	40	34.5	13	2 US-08-592-406-20	Sequence 20, Appl
32	40	34.5	119	1 US-08-340-539A-18	Sequence 18, Appl
33	40	34.5	280	1 US-08-434-255-8	Sequence 8, Appl
34	40	34.5	280	1 US-08-459-967-8	Sequence 8, Appl
35	40	34.5	280	1 US-08-460-327-8	Sequence 8, Appl
36	40	34.5	280	1 US-08-459-871-8	Sequence 8, Appl
37	40	34.5	280	4 US-09-024-532-2	Sequence 2, Appl
38	40	34.5	280	4 US-09-104-623A-2	Sequence 2, Appl
39	40	34.5	306	4 US-09-457-046B-4	Sequence 4, Appl
40	40	34.5	340	5 PCT-US96-02331-13	Sequence 13, Appl
41	40	34.5	370	1 US-08-434-255-6	Sequence 6, Appl
42	40	34.5	370	1 US-08-459-967-6	Sequence 6, Appl
43	40	34.5	370	1 US-08-460-327-6	Sequence 6, Appl
44	40	34.5	370	1 US-08-459-871-6	Sequence 6, Appl
45	40	34.5	397	1 US-08-434-255-2	Sequence 2, Appl

## ALIGNMENTS

```
RESULT 1
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match 100.0%; Score 116; DB 3; Length 1430;
Best Local Similarity 100.0%; Pred. No. 8.2e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AINHLSTLEAWSNDNDPOYNKD 21
    |||||
Db 495 AINHLSTLEAWSNDNDPOYNKD 515

RESULT 2
US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6281479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
```

EARLIER APPLICATION NUMBER: 09/008,172  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/482,711  
EARLIER FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 6  
LENGTH: 1430  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-210-361-6

Query Match 100.0%: Score 116; DB 4; Length 1430;  
Best Local Similarity 100.0%: Pred. No. 8,2e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AINHLSTLEAWSNDNDPOYNKD 21  
|||||  
495 AINHLSTLEAWSNDNDPOYNKD 515

RESULT 3  
US-08-793-824-2  
Sequence 2, Application US/08793824  
Patent No. 5981838  
GENERAL INFORMATION:  
APPLICANT: Simpson, Christine Lynn  
APPLICANT: Giffard, Phillip Morrison  
APPLICANT: Jacques, Nicholas Anthony  
TITLE OF INVENTION: Genetic Manipulation of Plants to  
TITLE OF INVENTION: Increase Stored Carbohydrates  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Griffith Hack & Co  
STREET: Level 8, 168 Walker Street  
CITY: No. 5981838ch Sydney  
STATE: New South Wales  
COUNTRY: Australia  
ZIP: 2060  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/793,824  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PM7643  
FILING DATE: 24-AUG-1994  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 61 2 957 5944  
TELEFAX: 61 2 957 6288  
TELEX: 26547  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1577 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus salivarius  
US-08-793-824-2

Query Match 77.6%: Score 90; DB 2; Length 1577;  
Best Local Similarity 85.7%: Pred. No. 1.9e-06;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AINHLSTLEAWSNDNDPOYNKD 21

|||||  
DB 588 AIAHSLTLEAWSNDNDPOYNKD 608

RESULT 4  
US-09-007-999-2  
Sequence 2, Application US/09007999  
Patent No. 6087559  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Substitutes for Modified Starch and  
FILE REFERENCE: 0356D  
CURRENT APPLICATION NUMBER: US/09/007,999  
CURRENT FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/478,704  
EARLIER FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1475  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-007-999-2

Query Match 64.7%: Score 75; DB 3; Length 1475;  
Best Local Similarity 71.4%: Pred. No. 0.00053;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 AINHLSTLEAWSNDNDPOYNKD 21  
|:|||||  
DB 481 ANDHSLTLEAWSNDNDPTYLHD 501

RESULT 5  
US-09-210-361-2  
Sequence 2, Application US/09210361  
Patent No. 6284479  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Substitutes for Modified Starches and  
FILE REFERENCE: 0357CR  
CURRENT APPLICATION NUMBER: US/09/210,361  
CURRENT FILING DATE: 1998-12-11  
EARLIER APPLICATION NUMBER: 09/007,999  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/478,704  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 09/009,620  
EARLIER FILING DATE: 1998-01-20  
EARLIER APPLICATION NUMBER: 08/485,243  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 09/008,172  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/482,711  
EARLIER FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1475  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-210-361-2

Query Match 64.7%: Score 75; DB 4; Length 1475;  
Best Local Similarity 71.4%: Pred. No. 0.00053;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 AINHLSTLEAWSNDNDPOYNKD 21  
|:|||||

Db 481 ANDHLSILEAMSDNDPYLHD 501

## RESULT 6

US-09-210-361-4  
; Sequence 4, Application US/09210361  
; Patent No. 6284479

## GENERAL INFORMATION:

APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Substitutes for Modified Starches and  
TITLE OF INVENTION: Latexes in Paper Manufacture  
FILE REFERENCE: 0357CR  
CURRENT APPLICATION NUMBER: US/09/210.361  
CURRENT FILING DATE: 1998-12-11  
EARLIER APPLICATION NUMBER: 09/007,999  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/478,704  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 09/009,620  
EARLIER FILING DATE: 1998-01-20  
EARLIER APPLICATION NUMBER: 08/485,243  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 09/008,172  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/482,711  
EARLIER FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 1375  
TYPE: PRT  
ORGANISM: streptococcus mutans  
US-09-210-361-4

Query Match 56.9%; Score 66; DB 4; Length 1375;  
Best Local Similarity 66.7%; Pred. No. 0.015;  
Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Oy 1 AINHLSILEAMSDNDPYNKD 21

Db 507 ANDHLSILEAMSYNDPYLHD 527

## RESULT 7

US-09-422-869-27  
; Sequence 27, Application US/09422869  
; Patent No. 6235481

## GENERAL INFORMATION:

APPLICANT: POLONSKY, KENNETH S.  
APPLICANT: HORIKAWA, YUKIO  
APPLICANT: ODA, NAOHISA  
APPLICANT: COX, NANCY J.  
APPLICANT: SREENAN, SEAMUS  
APPLICANT: ZHOU, YUN-PING  
APPLICANT: OTANI, KENICHI  
APPLICANT: HANIS, CRAIG L.  
APPLICANT: BELL, GRAEME I.  
TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES  
FILE REFERENCE: ARCD:307  
CURRENT APPLICATION NUMBER: US/09/422,869  
CURRENT FILING DATE: 1999-10-21  
EARLIER APPLICATION NUMBER: 60/134,175  
EARLIER FILING DATE: 1999-05-13  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 27  
LENGTH: 703  
TYPE: PRT  
ORGANISM: RAT  
US-09-422-869-27

Query Match 39.7%; Score 46; DB 4; Length 703;  
Best Local Similarity 70.0%; Pred. No. 14;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 10 AMSDNDPOYN 19

Db 296 AMSDNAPEWN 305

## RESULT 8

US-08-508-761B-31  
; Sequence 31, Application US/08508761B  
; Patent No. 6027520

## GENERAL INFORMATION:

APPLICANT: Jolliff, Gwennael  
APPLICANT: Guyonvarch, Armel  
APPLICANT: Purification, Relano  
APPLICANT: Duchiron, Francis  
APPLICANT: Renaud, Michel  
TITLE OF INVENTION: System for Protein Expression and  
TITLE OF INVENTION: Secretion Especially in Corynebacteria  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Jacobson, Price, Holman & Stern, PLLC  
STREET: 400 Seventh St. N.W.  
CITY: Washington D.C.  
COUNTRY: U.S.A.  
ZIP: 20004

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/508,761B  
FILING DATE: 31-JUL-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 91/09652  
FILING DATE: 29-JUL-1991

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 91/09870  
FILING DATE: 02-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Player, William E.  
REGISTRATION NUMBER: 31,409

## REFERENCE/DOCKET NUMBER: P58525NA

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 638-6666  
TELEFAX: (202) 393-5350  
INFORMATION FOR SEQ ID NO: 31:

## SEQUENCE CHARACTERISTICS:

LENGTH: 324 amino acids  
TYPE: amino acid  
STRANDEDNESS:

TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:

ORGANISM: C. glutamicum  
US-08-508-761B-31

Query Match 38.8%; Score 45; DB 3; Length 324;  
Best Local Similarity 53.8%; Pred. No. 8.2;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Oy 9 EAMSDNDPOYNKD 21

Db 212 ERWQENDPKSND 224

## RESULT 9

US-08-508-761B-2  
; Sequence 2, Application US/08508761B

Patent No. 5027920  
GENERAL INFORMATION:  
APPLICANT: Joliff, Gwennael  
APPLICANT: Guyonvatch, Arnel  
APPLICANT: Purification, Relano  
APPLICANT: Duchiron, Francis  
APPLICANT: Renaud, Michel  
TITLE OF INVENTION: System for Protein Expression and  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jacobson, Price, Holman & Stern, PLLC  
STREET: 400 Seventh St. N.W.  
CITY: Washington D.C.  
COUNTRY: U.S.A.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/508,761B  
FILING DATE: 31-JUL-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 91/09652  
FILING DATE: 29-JUL-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 91/09870  
FILING DATE: 02-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Player, William E.  
REGISTRATION NUMBER: 31,409  
REFERENCE/DOCKET NUMBER: P85525NA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 638-6666  
TELEFAX: (202) 393-5350  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 657 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
IS-08-508-761b-2

```

Query March 31, 1980
Best Local Similarity 53.8%; Score 45; DB 3; Length 657;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 9 EAWSNDPQYKND 21
      | | : | | | : | |
Db 286 ERWQNDPKSNVD 298

RESULT 10
US-08-277-231A-3
Sequence 31, Application US/08277231A
Patent No. 5643725

GENERAL INFORMATION:
APPLICANT: Green, Bruce A.
APPLICANT: Brinton, Charles C.
TITLE OF INVENTION: Sequence and Analysis of LKP P111n
Patent No. 5643725
TITLE OF INVENTION: Structural Genes and The LKP P111 Operon of No. 5643725
TITLE OF INVENTION: Haemophilus Influenzae
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.

```

```

1 ZIP: 02173
2
3 COMPUTER READABLE FORM:
4
5 MEDIUM TYPE: Floppy disk
6
7 COMPUTER: IBM PC compatible
8
9 OPERATING SYSTEM: PC-DOS/MS-DOS
10
11 SOFTWARE: PatentIn Release #1.0, Version #1.30
12
13 CURRENT APPLICATION DATA:
14
15 APPLICATION NUMBER: US/08/277,231A
16
17 FILING DATE: 19-JUL-1994
18
19 CLASSIFICATION: 435
20
21 ATTORNEY/AGENT INFORMATION:
22
23 NAME: Carroll, Alice O.
24
25 REGISTRATION NUMBER: 32,542
26
27 REFERENCE/DOCKET NUMBER: ACC94-02
28
29 TELECOMMUNICATION INFORMATION:
30
31 TELEPHONE: (617) 861-6240
32
33 TELEFAX: (617) 861-9540
34
35 INFORMATION FOR SEQ ID NO: 3:
36
37 SEQUENCE CHARACTERISTICS:
38
39 LENGTH: 259 amino acids
40
41 type: amino acid
42
43 TOPOLOGY: linear
44
45 MOLECULE TYPE: protein
46
47
48 US-08-277-231A-3

```

Query Match	38.4%	Score 44.5	DB 1	Length 259
Best Local Similarity	40.0%	Pred. No. 7.6		
Matches	8	Conservative	5	Mismatches 2
				Indels 5
				Gaps 1
QY	6	SILEAMSDN-----DPQYNK	20	
	::::			
Db	76	ALVCAQAMIDNGNPNADPKRYTK	95	

RESULT 11  
 US-08-473-750-6  
 . Sequence 6, Application US/08473750  
 . Patent No. 5834187 5786143  
 . Patent No. 5834187 5786143  
 . GENERAL INFORMATION:  
 . APPLICANT: Green, Bruce A.  
 . APPLICANT: Brinton, Jr., Charles C.  
 . TITLE OF INVENTION: Sequence and Analysis of LKP P11n  
 . Patent No. 5834187  
 . Patent No. 5834187 5786143  
 . TITLE OF INVENTION: Structural Gene and the LKP P11 Operon of No. 5834187 5786143  
 . TITLE OF INVENTION: Haemophilus Influenzae  
 . NUMBER OF SEQUENCES: 21  
 . CORRESPONDENCE ADDRESS:  
 . ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
 . STREET: Two Militia Drive  
 . CITY: Lexington  
 . STATE: Massachusetts  
 . COUNTRY: US  
 . ZIP: 02173  
 . COMPUTER READABLE FORM:  
 . MEDIUM TYPE: Floppy disk  
 . COMPUTER: IBM PC compatible  
 . OPERATING SYSTEM: PC-DOS/MS-DOS  
 . SOFTWARE: Patentin Release #1.0, Version #1.30  
 . CURRENT APPLICATION DATA:  
 . APPLICATION NUMBER: US/08/473,750  
 . FILING DATE: 07-JUN-1995  
 . CLASSIFICATION: 435  
 . PRIOR APPLICATION DATA:  
 . APPLICATION NUMBER: US 08/277,321  
 . FILING DATE: 19-JUL-1994  
 . ATTORNEY/AGENT INFORMATION:  
 . NAME: Carroll, Alice O.  
 . REGISTRATION NUMBER: 33,542  
 . REFERENCE/DOCKET NUMBER: ACC94-02B  
 . TELECOMMUNICATION INFORMATION:

```
QY      6 SILEAWSND-----DPQYNK 20
      ::::| | | | | | | |
```

CITY: NEW YORK  
STATE: New York  
COUNTRY: U.S.A.

ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/684,024  
FILING DATE: 19-JUL-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 1747/46621-A  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 205 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-684-024-6

Query Match 36.2% Score 42; DB 2; Length 205;  
Best Local Similarity 40.0%; Pred. NO. 15;  
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
OY 4 HSLILEANSNDNDPOY 18  
||:|::||:  
DB 160 HLVEPEKWEESGPQF 174

RESULT 15  
US-09-145-868-6  
Sequence 6, Application US/09145868  
Patent No. 6096522  
GENERAL INFORMATION:  
APPLICANT: Benzeira, Robert  
TITLE OF INVENTION: GENE ENCODING THE HUMAN HOMOLOG OF MAD2  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/145,868  
FILING DATE: 02-SEP-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 1747/46621-B  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 205 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-145-868-6

Query Match 36.2% Score 42; DB 3; Length 205;  
Best Local Similarity 40.0%; Pred. NO. 15;  
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
OY 4 HSLILEANSNDNDPOY 18  
||:|::||:  
DB 160 HLVEPEKWEESGPQF 174

Search completed: March 27, 2002, 13:59:30  
Job time: 583 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:20:24 ; Search time 1139.61 seconds

(without alignments)  
5.116 Million cell updates/sec

Title: US-09-290-049a-11

Perfect score: 116

Sequence: 1 AINHLSTLEAWSNDNPQYKND 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

Maximum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/2/paa/PCRUS\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US08\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep:\*  
8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep:\*  
9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep:\*  
10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep:\*  
11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep:\*  
12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep:\*  
13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep:\*  
14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep:\*  
15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep:\*  
16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep:\*  
17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep:\*  
18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep:\*  
19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep:\*  
20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep:\*  
21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep:\*  
22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep:\*  
23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep:\*  
24: /cgn2\_6/ptodata/2/paa/US60\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	116	100.0	21	US-09-290-049-11	Sequence 11, Appl
2	116	100.0	1430	US-09-649-885-2	Sequence 2, Appl
3	116	100.0	1430	US-09-740-274-6	Sequence 6, Appl
4	75	64.7	21	US-09-290-049-1	Sequence 1, Appl
5	75	64.7	21	US-09-290-049-10	Sequence 10, Appl
6	75	64.7	21	US-09-290-049-13	Sequence 13, Appl
7	75	64.7	1475	US-09-557-848-2	Sequence 2, Appl
8	75	64.7	1475	US-09-740-274-2	Sequence 2, Appl
9	74	63.8	21	US-09-290-049-12	Sequence 12, Appl

10	74	63.8	21	US-09-290-049-14	Sequence 14, Appl
11	66	56.9	1375	US-09-740-274-4	Sequence 4, Appl
12	64	55.2	2057	US-09-499-203-2	Sequence 2, Appl
13	49	42.2	623	PCT-US01-05614-482	Sequence 482, App
14	49	42.2	623	PCT-US98-11422A-407	Sequence 407, App
15	49	42.2	623	US-09-205-258-482	Sequence 482, App
16	49	42.2	623	US-09-333-767-482	Sequence 482, App
17	49	42.2	704	PCT-US01-05614-422	Sequence 422, App
18	49	42.2	704	US-09-205-258-422	Sequence 422, App
19	49	42.2	704	US-09-933-767-422	Sequence 422, App
20	48	41.4	883	US-09-489-039A-9716	Sequence 9716, App
21	47	40.5	196	US-09-595-298A-765	Sequence 765, App
22	47	40.5	521	US-09-107-532-6431	Sequence 6431, App
23	47	40.5	521	US-09-107-532A-6431	Sequence 6431, App
24	46.5	40.1	327	US-09-107-532-6181	Sequence 6181, App
25	46.5	40.1	327	US-09-107-532A-6181	Sequence 6181, App
26	46	39.7	199	US-09-391-631-2917	Sequence 2917, App
27	46	39.7	247	US-09-391-631-2916	Sequence 2916, App
28	46	39.7	647	US-09-281-253-10	Sequence 10, Appl
29	46	39.7	647	US-09-281-253-10	Sequence 10, Appl
30	46	39.7	647	US-09-281-253-10	Sequence 10, Appl
31	46	39.7	703	US-09-768-877-27	Sequence 10, Appl
32	46	39.7	1652	US-09-768-877-27	Sequence 10, Appl
33	46	39.7	10182	US-09-242-679-1100	Sequence 1100, App
34	46	39.7	10182	US-09-134-001C-3159	Sequence 3159, App
35	45	38.8	151	US-09-450-969-4098	Sequence 4098, App
36	45	38.8	205	US-09-198-452A-815	Sequence 815, App
37	45	38.8	205	US-09-438-185-767	Sequence 767, App
38	45	38.8	365	US-09-312-544-8416	Sequence 8416, App
39	45	38.8	657	US-09-602-839A-534	Sequence 534, App
40	45	38.8	657	US-08-039-028A-2	Sequence 2, Appl
41	45	38.8	657	US-08-039-028B-2	Sequence 2, Appl
42	45	38.8	657	US-08-039-028B-2	Sequence 2, Appl
43	45	38.8	657	US-08-039-028B-2	Sequence 2, Appl
44	45	38.8	790	US-09-738-626-6670	Sequence 6670, App
45	44.5	38.4	259	US-09-543-681A-6460	Sequence 6460, App
			6	US-08-277-231-3	Sequence 3, Appl

#### ALIGNMENTS

RESULT 1  
US-09-290-049-11  
Sequence 11, Application US/09290049  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
FILE REFERENCE: PDC98-0192A  
CURRENT FILING DATE: 1999-04-12  
EARLIER FILING DATE: 1998-04-13  
EARLIER APPLICATION NUMBER: 60/081,550  
EARLIER FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 11  
LENGTH: 21  
TYPE: PRT  
ORGANISM: S. mutans  
US-09-290-049-11

Query Match 100.0%; Score 116; DB 16; Length 21;  
Best Local Similarity 100.0%; Pred. No. 4e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 AINHLSTLEAWSNDNPQYKND 21  
Db 1 AINHLSTLEAWSNDNPQYKND 21

## RESULT 2

US-09-649-885-2  
; Sequence 2, Application US/09649885  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; TITLE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0358D2  
; CURRENT APPLICATION NUMBER: US/09/649,885  
; CURRENT FILING DATE: 2000-08-28  
; PRIOR APPLICATION NUMBER: US 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-649-885-2

Query Match 100.0%; Score 116; DB 20; Length 1430;  
Best Local Similarity 100.0%; Pred. No. 4,4e-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AINHLSTLEAWSNDNDPQYKND 21  
| : |||||  
Db 495 AINHLSTLEAWSNDNDPQYKND 515

## RESULT 3

US-09-740-274-6  
; Sequence 6, Application US/09740274  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-740-274-6

Query Match 100.0%; Score 116; DB 21; Length 1430;  
Best Local Similarity 100.0%; Pred. No. 4,4e-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AINHLSTLEAWSNDNDPQYKND 21  
| : |||||  
Db 495 AINHLSTLEAWSNDNDPQYKND 515

## RESULT 4

US-09-290-049-1  
; Sequence 1, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARIES  
; FILE REFERENCE: FDC98-01P2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; CURRENT FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: EAW peptide  
US-09-290-049-1

Query Match 64.7%; Score 75; DB 16; Length 21;  
Best Local Similarity 71.4%; Pred. No. 0.00016;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 AINHLSTLEAWSNDNDPQYKND 21  
| : |||||  
Db 1 ANDHLSTLEAWSNDNDTPYLHD 21

## RESULT 5

US-09-290-049-10  
; Sequence 10, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: FDC98-01P2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; CURRENT FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 10  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: S. mutans  
US-09-290-049-10

Query Match 64.7%; Score 75; DB 16; Length 21;  
Best Local Similarity 71.4%; Pred. No. 0.00016;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 AINHLSTLEAWSNDNDPQYKND 21  
| : |||||  
Db 1 ANDHLSTLEAWSNDNDTPYLHD 21

RESULT 6  
US-09-290-049-13  
; Sequence 13, Application US/09290049  
; GENERAL INFORMATION:

APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
TITLE OF INVENTION: CARIES  
FILE REFERENCE: FDC98-01D2A  
CURRENT APPLICATION NUMBER: US/09/290,049  
CURRENT FILING DATE: 1999-04-12  
PRIOR APPLICATION NUMBER: 60/081,550  
EARLIER FILING DATE: 1998-04-13  
EARLIER APPLICATION NUMBER: 60/115,142  
EARLIER FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 13  
LENGTH: 21  
TYPE: PRT  
ORGANISM: S. downei  
US-09-290-049-13

Query Match 64.7%; Score 75; DB 16; Length 21;  
Best Local Similarity 75.0%; Pred. No. 0.00016;  
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 AINHLSTLEAWSNDNDPOYNK 20  
DB 1 AIDHLSILEAWSNDNDPYLHD 20

RESULT 7  
US-09-557-848-2  
Sequence 2, Application US/09557848  
GENERAL INFORMATION:  
APPLICANT: Nicholas, Scott E.  
TITLE OF INVENTION: Substitutes for Modified Starch and  
FILE REFERENCE: 0356D2  
CURRENT APPLICATION NUMBER: US/09/557,848  
CURRENT FILING DATE: 2000-04-26  
EARLIER APPLICATION NUMBER: 08/478,704  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 09/007,999  
EARLIER FILING DATE: 1998-01-16  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1475  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-557-848-2

Query Match 64.7%; Score 75; DB 19; Length 1475;  
Best Local Similarity 71.4%; Pred. No. 0.018;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 AINHLSTLEAWSNDNDPOYNK 21  
DB 481 ANDHLSILEAWSNDNDPYLHD 501

RESULT 8  
US-09-740-274-2  
Sequence 2, Application US/09740274  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Glucan-containing Compositions and Paper  
FILE REFERENCE: 0357CRD  
CURRENT APPLICATION NUMBER: US/09/740,274  
CURRENT FILING DATE: 2000-12-19  
PRIOR APPLICATION NUMBER: 09/210,361  
PRIOR FILING DATE: 1998-12-11  
PRIOR APPLICATION NUMBER: 09/007,999

PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/478,704  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/009,620  
PRIOR FILING DATE: 1998-01-20  
PRIOR APPLICATION NUMBER: 08/485,243  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/008,172  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/482,711  
PRIOR FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1475  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-740-274-2

Query Match 64.7%; Score 75; DB 21; Length 1475;  
Best Local Similarity 71.4%; Pred. No. 0.018;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 AINHLSTLEAWSNDNDPOYNK 21  
DB 481 ANDHLSILEAWSNDNDPYLHD 501

RESULT 9  
US-09-290-049-12  
Sequence 12, Application US/09290049  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
TITLE OF INVENTION: CARIES  
FILE REFERENCE: FDC98-01D2A  
CURRENT APPLICATION NUMBER: US/09/290,049  
CURRENT FILING DATE: 1999-04-12  
EARLIER APPLICATION NUMBER: 60/081,550  
EARLIER FILING DATE: 1998-04-13  
EARLIER APPLICATION NUMBER: 60/115,142  
EARLIER FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 12  
LENGTH: 21  
TYPE: PRT  
ORGANISM: S. downei  
US-09-290-049-12

Query Match 63.8%; Score 74; DB 16; Length 21;  
Best Local Similarity 66.7%; Pred. No. 0.00023;  
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 AINHLSTLEAWSNDNDPOYNK 21  
DB 1 ANNHVSIVEAWSNDNDPYLHD 21

RESULT 10  
US-09-290-049-14  
Sequence 14, Application US/09290049  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
TITLE OF INVENTION: CARIES  
FILE REFERENCE: FDC98-01D2A  
CURRENT APPLICATION NUMBER: US/09/290,049  
CURRENT FILING DATE: 1999-04-12

EARLIER APPLICATION NUMBER: 60/081,550  
EARLIER FILING DATE: 1998-04-13  
EARLIER APPLICATION NUMBER: 60/115,142  
EARLIER FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 14  
LENGTH: 21  
TYPE: PRT  
ORGANISM: S. sobrinus  
US-09-290-049-14

Query Match 63.8%; Score 74; DB 16; Length 21;  
Best Local Similarity 66.7%; Pred. No. 0.00023;  
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 AINHLSTLEWMSNDPQYKND 21  
1 ANNHVSIVEWMSNDPQYKND 21

RESULT 11  
US-09-740-274-4  
Sequence 4, Application US/09740274  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Glucan-containing Compositions and Paper  
FILE REFERENCE: 0357CARD  
CURRENT APPLICATION NUMBER: US/09/740,274  
CURRENT FILING DATE: 2000-12-19  
PRIOR APPLICATION NUMBER: 09/210,361  
PRIOR FILING DATE: 1998-12-11  
PRIOR APPLICATION NUMBER: 09/007,999  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/478,704  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/009,620  
PRIOR FILING DATE: 1998-01-20  
PRIOR APPLICATION NUMBER: 08/485,243  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/008,172  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/482,711  
PRIOR FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 1375  
TYPE: PRT  
ORGANISM: streptococcus mutans  
US-09-740-274-4

Query Match 56.9%; Score 66; DB 21; Length 1375;  
Best Local Similarity 66.7%; Pred. No. 0.46;  
Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 AINHLSTLEWMSNDPQYKND 21  
1 ANNHVSIVEWMSNDPQYKND 21

Db 507 ANDHLSTLEWMSNDPQYKND 527

RESULT 12  
US-09-499-203-2  
Sequence 2, Application US/09499203  
GENERAL INFORMATION:  
APPLICANT: KOSSMANN, Jens  
APPLICANT: WELSH, Thomas  
APPLICANT: QUANZ, Martin  
APPLICANT: KNUTH, Karola  
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase  
FILE REFERENCE: 147-196P

CURRENT APPLICATION NUMBER: US/09/499,203  
CURRENT FILING DATE: 2000-02-08  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 2057  
TYPE: PRT  
ORGANISM: Leuconostoc mesenteroides  
US-09-499-203-2

Query Match 55.2%; Score 64; DB 18; Length 2057;  
Best Local Similarity 66.7%; Pred. No. 1.5;  
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 AINHLSTLEWMSNDPQY 18  
1 ANNHVSIVEWMSNDPQY 18

Db 665 ANHLSTLEWMSNDPQY 682

RESULT 13  
PCT-US01-05614-482  
Sequence 482, Application PC/TUS0105614  
GENERAL INFORMATION:  
APPLICANT: Human Genome Sciences, Inc.  
TITLE OF INVENTION: 207 Human Secreted Proteins  
FILE REFERENCE: P2007PCT2  
CURRENT APPLICATION NUMBER: PCT/US01/05614  
CURRENT FILING DATE: 2001-02-21  
PRIOR APPLICATION NUMBER: 60/184,836  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: 60/193,170  
PRIOR FILING DATE: 2000-03-29  
NUMBER OF SEQ ID NOS: 1245  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 482  
LENGTH: 623  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (111)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (575)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
PCT-US01-05614-482

Query Match 42.2%; Score 49; DB 1; Length 623;  
Best Local Similarity 52.9%; Pred. No. 1e+02;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 LSTLEWMSNDPQYKND 21  
1 ANNHVSIVEWMSNDPQYKND 21

Db 107 IWLXGASEFDPQYKND 123

RESULT 14  
PCT-US98-11422A-407  
Sequence 407, Application PC/TUS9811422A  
GENERAL INFORMATION:  
APPLICANT: Human Genome Sciences, Inc., et al.  
TITLE OF INVENTION: 207 Human Secreted Proteins  
NUMBER OF SEQUENCES: 800  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.,  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:



Db 107 IMVLXGASEFDPQYNKD 123

Search completed: March 27, 2002, 14:20:25  
Job time: 1572 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:22:47 ; Search time 137.48 Seconds

(without alignments)  
10,540 Million cell updates/sec

Title: US-09-290-049A-11

Perfect score: 116

Sequence: 1 AINHLSTLEAMSDNDPOYKND 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 332938 seqs, 6899538 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PCF\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep1:\*  
7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep1:\*  
8: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	116	100.0	21	6	US-09-562-328-22
2	116	100.0	21	6	US-09-290-049A-11
3	116	100.0	545	6	US-09-604-957-4
4	75	64.7	21	6	US-09-562-328-20
5	75	64.7	21	6	US-09-562-328-24
6	75	64.7	21	6	US-09-290-049A-1
7	75	64.7	21	6	US-09-290-049A-10
8	75	64.7	21	6	US-09-290-049A-13
9	74	63.8	21	6	US-09-562-328-23
10	74	63.8	21	6	US-09-562-328-25
11	74	63.8	21	6	US-09-290-049A-12
12	74	63.8	21	6	US-09-290-049A-14
13	68	58.6	523	6	US-09-604-957-5
14	66	56.9	21	6	US-09-562-328-21
15	64	55.2	584	6	US-09-604-957-6
16	53.5	46.1	535	6	US-09-604-957-7
17	53.5	46.1	1278	6	US-09-604-957-3
18	51	44.0	429	1	PCT-US02-03987-14022
19	51	44.0	429	6	US-09-815-242-14022
20	51	44.0	429	7	US-10-072-851-14022
21	49	42.2	623	7	US-10-023-282-482
22	49	42.2	704	7	US-10-023-282-422
23	49	42.2	704	6	US-09-611-526-3841
24	47	40.5	95	6	US-09-945-301-14
25	46	39.7	371	6	US-09-708-427-29185

26	46	39.7	372	6	US-09-708-427-29184	Sequence 29184, A
27	46	39.7	411	6	US-09-620-394B-1367	Sequence 1367, Ap
28	46	39.7	412	6	US-09-620-394B-1366	Sequence 1366, Ap
29	46	39.7	420	6	US-09-708-427-29183	Sequence 29183, A
30	46	39.7	460	6	US-09-620-394B-1365	Sequence 1365, Ap
31	45	38.8	330	6	US-09-708-427-72864	Sequence 72864, A
32	45	38.8	364	6	US-09-708-427-72863	Sequence 72863, A
33	45	38.8	376	6	US-09-708-427-72862	Sequence 72862, A
34	45	38.8	596	6	US-09-708-427-31640	Sequence 31640, A
35	45	38.8	597	6	US-09-708-427-31639	Sequence 31639, A
36	45	38.8	637	6	US-09-708-427-31638	Sequence 31638, A
37	44	37.9	869	6	US-09-614-150-423	Sequence 423, App
38	44	37.9	1017	6	US-09-614-150-35226	Sequence 35226, A
39	43.5	37.5	691	6	US-09-313-942-20	Sequence 20, Appl
40	43.5	37.5	694	6	US-09-313-942-18	Sequence 18, Appl
41	43.5	37.5	694	6	US-09-313-942-22	Sequence 22, Appl
42	43.5	37.5	784	6	US-09-313-942-30	Sequence 30, Appl
43	43.5	37.5	789	6	US-09-611-526-2651	Sequence 2651, Ap
44	43.5	37.5	793	6	US-09-313-942-32	Sequence 32, Appl
45	43.5	37.5	825	7	US-10-010-802-3	Sequence 3, Appl

## ALIGNMENTS

RESULT 1  
US-09-562-328-22  
Sequence 22, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARRIES  
FILE REFERENCE: 0495, 0046-01  
CURRENT FILING DATE: US/09/562,328  
CURRENT APPLICATION NUMBER: 2000-05-01  
PRIOR FILING DATE: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 22  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-22

Query Match 100.0%; Score 116; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 9.3e-12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AINHLSTLEAMSDNDPOYKND 21  
DB 1 AINHLSTLEAMSDNDPOYKND 21

RESULT 2  
US-09-290-049A-11  
Sequence 11, Application US/09290049A  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL CARRIES  
FILE REFERENCE: 1564, 1008-002  
CURRENT APPLICATION NUMBER: US/09/290,049A  
CURRENT FILING DATE: 1999-04-12  
PRIOR APPLICATION NUMBER: 60/081,550  
PRIOR FILING DATE: 1998-04-13  
PRIOR APPLICATION NUMBER: 60/115,142  
PRIOR FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 11  
LENGTH: 21  
TYPE: PRT  
ORGANISM: S. mutans  
US-09-290-049a-11

Query Match 100.0%; Score 116; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 9.3e-12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AINHLSLEAWSNDPQYKND 21  
|||||  
DB 1 AINHLSLEAWSNDPQYKND 21

RESULT 3  
US-09-604-957-4

Sequence 4, Application US/09604957  
GENERAL INFORMATION:

APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
APPLICANT: DIJKHUIZEN, LUBBERT  
APPLICANT: RAHAOUI, HAKIM  
APPLICANT: LEER, ROBERT-JAN  
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
FILE REFERENCE: BO 43388  
CURRENT APPLICATION NUMBER: US/09/604,957  
CURRENT FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 545  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-604-957-4

Query Match 100.0%; Score 116; DB 6; Length 545;  
Best Local Similarity 100.0%; Pred. No. 3.9e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AINHLSLEAWSNDPQYKND 21  
|||||  
DB 75 AINHLSLEAWSNDPQYKND 95

SULT 4

US-09-562-328-20  
Sequence 20, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
APPLICANT: SMITH, DANIEL J.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 04995.0046-01  
CURRENT APPLICATION NUMBER: US/09/562,328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 20  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-20

Query Match 64.7%; Score 75; DB 6; Length 21;  
Best Local Similarity 71.4%; Pred. No. 2.4e-05;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 AINHLSLEAWSNDPQYKND 21  
|:|||||  
DB 1 ANDHLSLEAWSNDPQYKND 21

RESULT 5  
US-09-562-328-24

Sequence 24, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
APPLICANT: SMITH, DANIEL J.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 04995.0046-01  
CURRENT APPLICATION NUMBER: US/09/562,328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 24  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-24

Query Match 64.7%; Score 75; DB 6; Length 21;  
Best Local Similarity 75.0%; Pred. No. 2.4e-05;  
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AINHLSLEAWSNDPQYKND 20  
|:|||||  
DB 1 AIDHLSLEAWSNDPQYKND 20

RESULT 6

US-09-290-049a-1  
Sequence 1, Application US/09290049a  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
FILE REFERENCE: 1564.1008-002  
CURRENT APPLICATION NUMBER: US/09/290,049a  
CURRENT FILING DATE: 1999-04-12  
PRIOR APPLICATION NUMBER: 60/081,550  
PRIOR FILING DATE: 1998-04-13  
PRIOR APPLICATION NUMBER: 60/115,142  
PRIOR FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: EAW peptide  
US-09-290-049a-1

Query Match 64.7%; Score 75; DB 6; Length 21;  
Best Local Similarity 71.4%; Pred. No. 2.4e-05;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 AINHLSLEAWSNDPQYKND 21  
|:|||||  
DB 1 ANDHLSLEAWSNDPQYKND 21

RESULT 7



```
US-09-290-049a-10
: Sequence 10, Application US/09290049A
: GENERAL INFORMATION:
: APPLICANT: Smith, Daniel J.
: APPLICANT: Taubman, Martin A.
: TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
: FILE REFERENCE: 1564.1008-002
: CURRENT APPLICATION NUMBER: US/09/290.049A
: CURRENT FILING DATE: 1999-04-12
: PRIOR APPLICATION NUMBER: 60/081,550
: PRIOR FILING DATE: 1998-04-13
: PRIOR APPLICATION NUMBER: 60/115,142
: PRIOR FILING DATE: 1999-01-08
: NUMBER OF SEQ ID NOS: 19
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 10
: LENGTH: 21
: TYPE: PRT
: ORGANISM: S. mutans
: US-290-049a-10

Query Match          64.7%; Score 75; DB 6; Length 21;
Best Local Similarity 71.4%; Pred. No. 2.4e-05;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 AINHSLIEAWSNDPQYKND 21
   1 AINHSLIEAWSNDPQYKND 21
   1 AINHSLIEAWSNDPQYKND 21

Db 1 AINHSLIEAWSNDPQYKND 21

RESULT 8
US-09-290-049a-13
: Sequence 13, Application US/09290049A
: GENERAL INFORMATION:
: APPLICANT: Smith, Daniel J.
: APPLICANT: Taubman, Martin A.
: TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
: FILE REFERENCE: 1564.1008-002
: CURRENT APPLICATION NUMBER: US/09/290.049A
: CURRENT FILING DATE: 1999-04-12
: PRIOR APPLICATION NUMBER: 60/081,550
: PRIOR FILING DATE: 1998-04-13
: PRIOR APPLICATION NUMBER: 60/115,142
: PRIOR FILING DATE: 1999-01-08
: NUMBER OF SEQ ID NOS: 19
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 13
: LENGTH: 21
: TYPE: PRT
: ORGANISM: S. downei
: US-09-290-049a-13

Query Match          64.7%; Score 75; DB 6; Length 21;
Best Local Similarity 75.0%; Pred. No. 2.4e-05;
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 AINHSLIEAWSNDPQYKND 20
   1 AINHSLIEAWSNDPQYKND 20
   1 AINHSLIEAWSNDPQYKND 20

Db 1 AINHSLIEAWSNDPQYKND 20

RESULT 9
US-09-562-328-23
: Sequence 23, Application US/09562328
: GENERAL INFORMATION:
: APPLICANT: LEES, ANDREW
: APPLICANT: TAUBMAN, MARTIN A.
: APPLICANT: SMITH, DANIEL J.
: TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
: FILE REFERENCE: 04995.0046-01
: CURRENT APPLICATION NUMBER: US/09/562.328
: CURRENT FILING DATE: 2000-05-01
: PRIOR APPLICATION NUMBER: 09/288,965
: PRIOR FILING DATE: 1999-04-09
: NUMBER OF SEQ ID NOS: 44
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 23
: LENGTH: 21
: TYPE: PRT
: ORGANISM: Streptococcus sp.
: US-09-562-328-23

Query Match          63.8%; Score 74; DB 6; Length 21;
Best Local Similarity 66.7%; Pred. No. 3.4e-05;
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 AINHSLIEAWSNDPQYKND 21
   1 AINHSLIEAWSNDPQYKND 21
   1 AINHSLIEAWSNDPQYKND 21

Db 1 AINHSLIEAWSNDPQYKND 21

RESULT 10
US-09-562-328-25
: Sequence 25, Application US/09562328
: GENERAL INFORMATION:
: APPLICANT: LEES, ANDREW
: APPLICANT: TAUBMAN, MARTIN A.
: APPLICANT: SMITH, DANIEL J.
: TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
: FILE REFERENCE: 04995.0046-01
: CURRENT APPLICATION NUMBER: US/09/562.328
: CURRENT FILING DATE: 2000-05-01
: PRIOR APPLICATION NUMBER: 09/288,965
: PRIOR FILING DATE: 1999-04-09
: NUMBER OF SEQ ID NOS: 44
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 25
: LENGTH: 21
: TYPE: PRT
: ORGANISM: Streptococcus sp.
: US-09-562-328-25

Query Match          63.8%; Score 74; DB 6; Length 21;
Best Local Similarity 66.7%; Pred. No. 3.4e-05;
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 AINHSLIEAWSNDPQYKND 21
   1 AINHSLIEAWSNDPQYKND 21
   1 AINHSLIEAWSNDPQYKND 21

Db 1 AINHSLIEAWSNDPQYKND 21

RESULT 11
US-09-290-049a-12
: Sequence 12, Application US/09290049A
: GENERAL INFORMATION:
: APPLICANT: Smith, Daniel J.
: APPLICANT: Taubman, Martin A.
: TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
: FILE REFERENCE: 1564.1008-002
: CURRENT APPLICATION NUMBER: US/09/290.049A
: CURRENT FILING DATE: 1999-04-12
: PRIOR APPLICATION NUMBER: 60/081,550
: PRIOR FILING DATE: 1998-04-13
: PRIOR APPLICATION NUMBER: 60/115,142
: PRIOR FILING DATE: 1999-01-08
: NUMBER OF SEQ ID NOS: 19
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 12
: LENGTH: 21
```

```
FILE REFERENCE: 04995.0046-01
: CURRENT APPLICATION NUMBER: US/09/562.328
: CURRENT FILING DATE: 2000-05-01
: PRIOR APPLICATION NUMBER: 09/288,965
: PRIOR FILING DATE: 1999-04-09
: NUMBER OF SEQ ID NOS: 44
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 23
: LENGTH: 21
: TYPE: PRT
: ORGANISM: Streptococcus sp.
: US-09-562-328-23

Query Match          63.8%; Score 74; DB 6; Length 21;
Best Local Similarity 66.7%; Pred. No. 3.4e-05;
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 AINHSLIEAWSNDPQYKND 21
   1 AINHSLIEAWSNDPQYKND 21
   1 AINHSLIEAWSNDPQYKND 21

Db 1 AINHSLIEAWSNDPQYKND 21

RESULT 10
US-09-562-328-25
: Sequence 25, Application US/09562328
: GENERAL INFORMATION:
: APPLICANT: LEES, ANDREW
: APPLICANT: TAUBMAN, MARTIN A.
: APPLICANT: SMITH, DANIEL J.
: TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
: FILE REFERENCE: 04995.0046-01
: CURRENT APPLICATION NUMBER: US/09/562.328
: CURRENT FILING DATE: 2000-05-01
: PRIOR APPLICATION NUMBER: 09/288,965
: PRIOR FILING DATE: 1999-04-09
: NUMBER OF SEQ ID NOS: 44
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 25
: LENGTH: 21
: TYPE: PRT
: ORGANISM: Streptococcus sp.
: US-09-562-328-25

Query Match          63.8%; Score 74; DB 6; Length 21;
Best Local Similarity 66.7%; Pred. No. 3.4e-05;
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 AINHSLIEAWSNDPQYKND 21
   1 AINHSLIEAWSNDPQYKND 21
   1 AINHSLIEAWSNDPQYKND 21

Db 1 AINHSLIEAWSNDPQYKND 21

RESULT 11
US-09-290-049a-12
: Sequence 12, Application US/09290049A
: GENERAL INFORMATION:
: APPLICANT: Smith, Daniel J.
: APPLICANT: Taubman, Martin A.
: TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
: FILE REFERENCE: 1564.1008-002
: CURRENT APPLICATION NUMBER: US/09/290.049A
: CURRENT FILING DATE: 1999-04-12
: PRIOR APPLICATION NUMBER: 60/081,550
: PRIOR FILING DATE: 1998-04-13
: PRIOR APPLICATION NUMBER: 60/115,142
: PRIOR FILING DATE: 1999-01-08
: NUMBER OF SEQ ID NOS: 19
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 12
: LENGTH: 21
```

TYPE: PRT  
ORGANISM: S. downei  
US-09-290-049A-12

Query Match 63.8%; Score 74; DB 6; Length 21;  
Best Local Similarity 66.7%; Pred. No. 3.4e-05;  
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 AINHSLIEAWSNDNDPOYNKD 21  
DB 1 ANNHVSIVEAWSNDNDPTPLHD 21

RESULT 12  
US-09-290-049A-14  
Sequence 14, Application US/09290049A  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
FILE REFERENCE: 1564.1008-002  
CURRENT APPLICATION NUMBER: US/09/290.049A  
CURRENT FILING DATE: 1999-04-12  
PRIOR APPLICATION NUMBER: 60/081,550  
PRIOR FILING DATE: 1998-04-13  
PRIOR APPLICATION NUMBER: 60/115,142  
PRIOR FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 14  
LENGTH: 21  
TYPE: PRT  
ORGANISM: S. sobrinus  
US-09-290-049A-14

Query Match 63.8%; Score 74; DB 6; Length 21;  
Best Local Similarity 66.7%; Pred. No. 3.4e-05;  
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 AINHSLIEAWSNDNDPOYNKD 21  
DB 1 ANNHVSIVEAWSNDNDPTPLHD 21

RESULT 13  
US-09-604-957-5  
Sequence 5, Application US/09604957  
GENERAL INFORMATION:  
APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
APPLICANT: DIKHUIZEN, LOBBERT  
APPLICANT: RAHAOUI, HAKIM  
APPLICANT: LEER, ROBERT-JAN  
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
FILE REFERENCE: BO 43388  
CURRENT APPLICATION NUMBER: US/09/604.957  
CURRENT FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 523  
TYPE: PRT  
ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-5

Query Match 58.6%; Score 68; DB 6; Length 523;  
Best Local Similarity 66.7%; Pred. No. 0.012;  
Matches 14; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 AINHSLIEAWSNDNDPOYNKD 21  
DB 75 ANOHLSTLEDSHNDPTLYTD 95

RESULT 14  
US-09-562-328-21  
Sequence 21, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
APPLICANT: SMITH, DANIEL J.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 04995.0046-01  
CURRENT APPLICATION NUMBER: US/09/562.328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 21  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-21

Query Match 56.9%; Score 66; DB 6; Length 21;  
Best Local Similarity 66.7%; Pred. No. 0.00061;  
Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 AINHSLIEAWSNDNDPOYNKD 21  
DB 1 ANDHSLIEAWSNDNDPTPLHD 21

RESULT 15  
US-09-604-957-6  
Sequence 6, Application US/09604957  
GENERAL INFORMATION:  
APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
APPLICANT: DIKHUIZEN, LOBBERT  
APPLICANT: RAHAOUI, HAKIM  
APPLICANT: LEER, ROBERT-JAN  
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
FILE REFERENCE: BO 43388  
CURRENT APPLICATION NUMBER: US/09/604.957  
CURRENT FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 6  
LENGTH: 584  
TYPE: PRT  
ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-6

Query Match 55.2%; Score 64; DB 6; Length 584;  
Best Local Similarity 66.7%; Pred. No. 0.057;  
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 AINHSLIEAWSNDNDPOYNKD 18  
DB 75 ANKHSLIEAWNGKNDPOYNKD 92

Search completed: March 27, 2002, 14:22:47  
Job time: 1694 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:01:18 ; Search time 102.51 Seconds  
(without alignments)  
15.605 Million cell updates/sec

Title: US-09-290-049a-11

Perfect score: 116  
Sequence: 1 AINHL5LEAWSNDNDPOYKND 21

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	116	100.0	1431	2 A45866	dextranucrase (EC
2	90	77.6	1577	2 T30858	glucosyltransferas
3	85	73.3	1449	2 T30857	glucosyltransferas
4	85	73.3	1449	2 T30552	glucosyltransferas
5	81	69.8	1365	2 A41483	glucosyltransferas
6	78	67.2	1508	2 T31098	probable dextran
7	75	64.7	1475	2 B33135	glfB protein precu
8	74	63.8	1592	2 A38175	glucosyltransferas
9	68	58.6	1518	2 A44811	glucosyltransferas
10	67	57.8	1599	2 S22737	glucosyltransferas
11	66	56.9	1375	2 JT0345	dextranucrase (EC
12	49	42.2	632	2 T46504	hypothetical prote
13	47	40.5	175	2 C86205	hypothetical prote
14	47	40.5	403	2 B70961	probable esterase
15	47	40.5	418	1 FOXRL2	sigma 2 protein -
16	47	40.5	445	2 H75360	cytochrome P450 -
17	47	40.5	491	2 A86824	sensor protein kin
18	46	39.7	418	1 FOXRL5	sigma 2 protein -
19	46	39.7	420	2 T05877	hypothetical prote
20	46	39.7	623	2 T35377	probable membrane
21	46	39.7	626	2 D70178	PTS system, fructo
22	46	39.7	703	2 A48764	calpain (EC 3.4.22
23	45.5	39.2	250	2 G72495	probable polysulf
24	45	38.8	194	2 H72037	conserved hypotet
25	45	38.8	194	2 C86586	hypothetical prote
26	45	38.8	344	2 T21604	hypothetical prote
27	45	38.8	440	2 C84265	adenylsuccinate s
28	45	38.8	504	2 S51942	punin 2 precursor
29	45	38.8	637	2 T00548	hypothetical prote

30	45	38.8	657	2 S25184	cspl protein - Cor
31	45	38.8	1475	2 T29809	hypothetical prote
32	45	38.8	2358	2 T39569	probable alpha-glu
33	45	38.8	2371	2 T43432	alpha-glucan synth
34	44.5	38.4	188	2 A75382	hypothetical prote
35	44.5	38.4	236	2 S54428	hypothetical prote
36	44.5	38.4	241	2 S24978	hypothetical prote
37	44	37.9	312	2 S67667	hypothetical prote
38	44	37.9	335	2 C85642	hypothetical prote
39	44	37.9	504	2 T24818	hypothetical prote
40	44	37.9	837	2 S54624	hypothetical prote
41	43.5	37.5	825	1 A60386	hypothetical prote
42	43	37.1	267	1 S71020	peptidoglycan-link
43	43	37.1	267	2 C81937	competence lipopro
44	43	37.1	267	2 H81167	competence lipopro
45	43	37.1	376	2 H86878	aminotransferase l

#### ALIGNMENTS

```
RESULT 1
A45866
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C:Accession: A45866
R:Honda, O.; Kato, C.; Kuramitsu, H. K.
J. Gen. Microbiol. 136, 2099-2105, 1990
A:Title: Nucleotide sequence of the Streptococcus mutans glfD gene encoding the gluco
A:Reference number: A45866; MUID:91100958
A:Accession: A45866
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1431 <HON>
A:Cross-references: GB:M29296
C:Superfamily: cpl repeat homology
C:Keywords: glycosyltransferase; hexosyltransferase
F:181-201/Domain: cpl repeat homology <CP1>
F:1127-1146/Domain: cpl repeat homology <CP2>
F:1192-1211/Domain: cpl repeat homology <CP3>
F:1257-1276/Domain: cpl repeat homology <CP4>
F:1277-1297/Domain: cpl repeat homology <CP5>
F:1321-1340/Domain: cpl repeat homology <CP6>
F:1341-1361/Domain: cpl repeat homology <CP6>
F:1365-1404/Domain: cpl repeat homology <CP7>

Query Match 100.0%; Score 116; DB 2; Length 1431;
Best local similarity 100.0%; Pred. No. 1.7e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AINHL5LEAWSNDNDPOYKND 21
Db 495 AINHL5LEAWSNDNDPOYKND 515

RESULT 2
T30858
glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30858
R:Stimpson, C.L.; Giffard, P.M.; Jacques, N.A.
Infect. Immun. 63, 609-621, 1995
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for
A:Reference number: 220909; MUID:95122197
A:Accession: T30858
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1577 <SIM>
A:Cross-references: EMBL:L35928; NID:g662380; PID:g662381; PIDN:AAC41A13.1
C:Genetics:
```



A:Reference number: A33128  
A:Accession: A33128  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-171,173-641,'N',643-1475 <SH2>  
A:Experimental source: strain GS-5  
C:Superfamily: cpl repeat homology  
F:1096-1115/Domain: cpl repeat homology <CP1>  
F:1224-1243/Domain: cpl repeat homology <CP2>  
F:1289-1308/Domain: cpl repeat homology <CP3>  
F:1354-1373/Domain: cpl repeat homology <CP4>  
F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match 64.7%; Score 75; DB 2; Length 1475;  
Best Local Similarity 71.4%; Pred. No. 0.0035;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

1 AINHSLILEAMSDNDTPYLNK 21  
|:||||||||| |  
481 ANHSLILEAMSDNDTPYLNH 501

RESULT 8  
A38175  
glucosyltransferase precursor - Streptococcus sobrinus  
C:Species: Streptococcus sobrinus  
C>Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 15-Oct-1999  
C:Accession: A38175  
R:Abou, H.; Mtsamura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.  
J. Bacteriol. 173, 989-996, 1991  
A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus  
A:Reference number: A38175; MUID:91123227  
A:Accession: A38175  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1592 <RBO>  
A:Cross-references: GB:D90213; NID:g217032; PIDN:BAAL4241.1; PID:d1014946; PID:g217033  
C:Superfamily: cpl repeat homology  
F:1093-1112/Domain: cpl repeat homology <CP1>  
F:1222-1241/Domain: cpl repeat homology <CP2>  
F:1287-1306/Domain: cpl repeat homology <CP3>  
F:1330-1351/Domain: cpl repeat homology <CP4>  
F:1352-1371/Domain: cpl repeat homology <CP5>  
F:1402-1420/Domain: cpl repeat homology <CP6>  
F:1465-1484/Domain: cpl repeat homology <CP7>  
F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 63.8%; Score 74; DB 2; Length 1592;  
Best Local Similarity 66.7%; Pred. No. 0.0054;  
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

1 AINHSLILEAMSDNDPOYNK 21  
|:||||||||| |  
477 ANHSLILEAMSDNDTPYLNH 497

RESULT 9  
A44811  
glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 15-Oct-1999  
C:Accession: A44811; S22726; S28809  
R:Gilford, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
J. Gen. Microbiol. 137, 2577-2593, 1991  
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase gen  
A:Reference number: A44811; MUID:92148377  
A:Accession: A44811  
A:Molecule type: DNA  
A:Residues: 1-1518 <GIF>  
A:Cross-references: EMBL:Z11873; NID:g47526; PIDN:CAAT77900.1; PID:g47527  
A:Note: sequence extracted from NCBI backbone (NCBIN:81050, NCBIR:81052)

C:Genetics:  
A:Gene: glfF  
C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:1307-1326/Domain: cpl repeat homology <CP4>

Query Match 58.6%; Score 68; DB 2; Length 1518;  
Best Local Similarity 63.2%; Pred. No. 0.043;  
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

1 AINHSLILEAMSDNDPOYN 19  
|:||||||||| |  
501 ALAHSLILEAMSLNDNHYN 519

RESULT 10  
S22737  
glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C>Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000  
C:Accession: S22737; S28810; B44811; S22727  
R:Jacques, N.  
submitted to the EMBL Data Library, March 1992  
A:Reference number: S22726  
A:Accession: S22737  
A:Molecule type: DNA  
A:Residues: 1-1599 <JAC>  
A:Cross-references: EMBL:Z11872; NID:g47530; PIDN:CAAT77898.1; PID:g47531  
R:Gilford, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
J. Gen. Microbiol. 137, 2577-2593, 1991  
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase  
A:Reference number: A44811; MUID:92148377  
A:Accession: S28810  
A:Molecule type: DNA  
A:Residues: 1-51 <GIF>  
A:Cross-references: EMBL:Z11873  
C:Genetics:  
A:Gene: glfK  
C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:1456-1475/Domain: cpl repeat homology <CPR>

Query Match 57.8%; Score 67; DB 2; Length 1599;  
Best Local Similarity 66.7%; Pred. No. 0.064;  
Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

1 AINHSLILEAMSDNDPOY 18  
|:||||||||| |  
491 ALAHSLILEAMSHNDPY 508

RESULT 11  
JT0345  
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)  
N:Alternate names: sucrose 6-glucosyltransferase  
C:Species: Streptococcus mutans  
C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 24-Sep-1999  
C:Accession: JT0345; C33135  
R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.  
Gene 69, 101-109, 1988  
A:Title: Sequence analysis of the glfC gene from Streptococcus mutans GS-5.  
A:Reference number: JT0345; MUID:89137980  
A:Accession: JT0345  
A:Molecule type: DNA  
A:Residues: 1-1375 <UED>  
A:Experimental source: GS-5  
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
J. Bacteriol. 169, 4263-4270, 1987  
A:Title: Sequence analysis of the glfB gene from Streptococcus mutans.  
A:Reference number: A33135; MUID:87308013

A:Accession: C33135  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-349 <SHI>  
A:Cross-references: GB:M17361  
C:Genetics:  
A:Gene: gtfC  
C:Function:  
A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans  
C:Superfamily: cpl repeat homology  
C:Keywords: duplication; glycosyltransferase; hexosyltransferase  
F:1-94/Domain: signal sequence #status predicted <SIG>  
F:35-1375/Product: glucosyltransferase #status predicted <MAT>  
F:1126-1145/Domain: cpl repeat homology <CP1>  
F:1253-1272/Domain: cpl repeat homology <CP2>  
F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 56.9%; Score 66; DB 2; Length 1375;  
Best Local Similarity 66.7%; Pred. No. 0.077;  
Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Oy 1 AINHLSEAMSDNDPOYNKD 21  
Db 507 ANDHLSEAMSYNDTPYLHD 527

## RESULT 12

hypothetical protein DKFZp434M179.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T46504  
R:Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, January 2000  
A:Reference number: Z23029  
A:Accession: T46504  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-632 <AAA>  
A:Cross-references: EMBL:AL137338  
A:Experimental source: adult testis; clone DKFZp434M179  
C:Genetics:  
A:Note: DKFZp434M179.1

Query Match 42.2%; Score 49; DB 2; Length 632;  
Best Local Similarity 52.9%; Pred. No. 13;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Oy 5 ISILEAMSDNDPOYNKD 21  
Db 119 IMVLGASFEFDPYKND 135

## RESULT 13

hypothetical protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: C86205  
R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chih, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hulzar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matli, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: C86205

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-175 <STOC>  
A:Cross-references: GB:AE005172; NID:98954041; PIDN:AAF82215.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 40.5%; Score 47; DB 2; Length 175;  
Best Local Similarity 46.2%; Pred. No. 6.3;  
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Oy 2 INHSLSEAMSDN 14  
Db 108 LNHQEVIAWSDH 120

## RESULT 14

B70961  
probable esterase - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: B70961  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon ; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A:Reference number: A70500; MUID:98295987  
A:Accession: B70961  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-403 <COL>  
A:Cross-references: GB:292669; GB:AL133456; NID:93242271; PIDN:CAB07015.1; PID:918715  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: lipC  
C:Superfamily: Mycobacterium tuberculosis probable esterase

Query Match 40.5%; Score 47; DB 2; Length 403;  
Best Local Similarity 50.0%; Pred. No. 16;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Oy 4 HSLSEAMSDNDPOYNKD 21  
Db 241 HLSALAGITANDPOYQAE 258

## RESULT 15

sigma 2 protein - reovirus type 1 (strain Lang)  
N:Alternate names: core protein  
C:Species: reovirus type 1  
A:Note: host Homo sapiens (man)  
C:Date: 31-Mar-1989 #sequence\_revision 30-Jun-1993 #text\_change 16-Jul-1999  
C:Accession: M1306; A29708  
R:Dermody, T.S.; Schiff, L.A.; Nibert, M.L.; Coombs, K.M.; Fields, B.N.  
J. Virol. 65, 5721-5731, 1991  
A:Title: The 52 gene nucleotide sequences of prototype strains of the three reovirus  
A:Reference number: M1306; MUID:92015462  
A:Accession: M1306  
A:Molecule type: genomic RNA  
A:Residues: 1-418 <DER>  
A:Cross-references: GB:S59098  
R:George, C.X.; Crowe, A.; Munemitsu, S.M.; Atwater, J.A.; Samuel, C.E.  
Biochem. Biophys. Res. Commun. 147, 1153-1161, 1987  
A:Title: Biosynthesis of reovirus-specified polypeptides. Molecular cDNA cloning and  
-2.  
A:Reference number: A29708; MUID:88024195  
A:Accession: A29708  
A:Molecule type: genomic RNA



A:Residues: 1-88,'V',90-204,'Y',206-322,'VATSCMWSKCDEW' <GEO>  
A:Cross-references: GB:M17598; NID:g333748; PIDN:AAA47278.1; PID:g333749  
C:Genetics:  
A:Map position: segment S2  
C:Superfamily: reovirus sigma 2 protein  
C:Keywords: core protein

Query Match	40:58;	Score 47;	DB 1;	Length 418;
Best Local Similarity	61.58;	Pred. NO. 17;		
Matches	8;	Conservative	2;	Mismatches 3; Indels 0; Gaps 0;
OY	8	LEAWSNDNDPOYNK	20	
	:	:		
Db	372	LEAWAREDDOYNQ	384	

Search completed: March 27, 2002, 14:01:19  
Job time: 481 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:57:54 ; Search time 198.55 Seconds  
(without alignments)  
7.834 Million cell updates/sec

Title: US-09-290-049A-10  
Perfect score: 119  
Sequence: 1 ANDHLSTLEAMSDNDPYLHD 21

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq\_1101.\*  
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	90.8	1592	14	AA19925
2	63	52.9	1577	17	AA19047
3	60	50.4	2057	21	AA19067
4	51	42.9	486	22	AA19067
5	50	42.0	305	21	AA19067
6	50	42.0	339	21	AA19067
7	48.5	40.8	401	22	AA19067
8	47.5	39.9	534	21	AA19067
9	47	39.5	195	21	AA19067
10	47	39.5	338	15	AA19067
11	46	38.7	118	20	AA19067

12	46	38.7	199	22	AA190817
13	46	38.7	336	20	AA193244
14	45	37.8	236	20	AA193993
15	45	37.8	385	12	AA19238
16	44	37.0	62	21	AA194863
17	44	37.0	161	21	AA194863
18	44	37.0	206	21	AA194863
19	44	37.0	220	22	AA194863
20	44	37.0	384	21	AA194863
21	43	36.1	126	21	AA194863
22	43	36.1	126	21	AA194863
23	43	36.1	170	21	AA194863
24	43	36.1	170	21	AA194863
25	43	36.1	193	21	AA194863
26	43	36.1	194	21	AA194863
27	43	36.1	327	22	AA194863
28	43	36.1	335	20	AA194863
29	43	36.1	335	20	AA194863
30	43	36.1	1195	20	AA194863
31	43	36.1	2466	16	AA194863
32	43	36.1	2466	19	AA194863
33	43	36.1	2466	21	AA194863
34	43	36.1	2485	21	AA194863
35	42.5	35.7	325	22	AA194863
36	42.5	35.7	390	22	AA194863
37	42	35.3	123	22	AA194863
38	42	35.3	174	21	AA194863
39	42	35.3	178	10	AA194863
40	42	35.3	185	21	AA194863
41	42	35.3	189	21	AA194863
42	42	35.3	236	20	AA194863
43	42	35.3	239	20	AA194863
44	42	35.3	259	17	AA194863
45	42	35.3	298	19	AA194863

#### ALIGNMENTS

RESULT 1	AA19925	standard; Protein; 1592 AA.
ID	AA19925	
XX	AA19925	
AC	AA19925	
XX	28-JUN-1993	(first entry)
DT		
XX		
DE	Glucosyltransferase I.	
XX		
KW	GT-1; Streptococcus; dental; caries.	
OS	Streptococcus sobrinus.	
XX		
PN	JP05023188-A.	
XX		
DR	WPI: 1993-079449/10.	
XX		
XX	N-PSDB: AA037760.	
PT	DNA sequence glucosyltransferase-I - comprises Streptococcus	
PT	sobrinus DNA sequence with at least one nucleotide added or	
XX	deleted	
PS	Claim 13; Page 15; 29pp; Japanese.	

CC The DNA sequence from *Streptococcus sobrinus* strain 6715 encodes  
 CC glucosyltransferase-I (and mutants). The DNA was obtd. by treating  
 CC *S. sobrinus* 6715 with mutanolysin, extracting the chromosomal DNA,  
 CC partially digesting with SauAI and fractionating on agarose gel.  
 CC The 3-5 kbp fragment was ligated into pUC18 and *E. coli* JM109  
 CC transformed with it. A GT-1 expressing clone was isolated and  
 CC sequenced. The clone may be used in the development of a drug for  
 CC dental caries.  
 CC  
 XX Sequence 1592 AA:

Query Match 90.8%; Score 108; DB 14; Length 1592;  
 Best Local Similarity 85.7%; Pred. No. 1.2e-07;  
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHLSLEAWSNDPPLYLHD 21  
 |||:|||||  
 477 annhslseawsnddpylhd 497

## RESULT 2

AAR91047  
 ID AAR91047 standard; Protein; 1577 AA.

AC AAR91047;

DT 22-MAY-1996 (first entry)

DE Alpha-D-glucosyltransferase.

XX Alpha-D-glucosyltransferase; primer-independent; soluble glucan;  
 KM sucrose; transgenic plant; cloning; *Escherichia coli*;  
 KM phage lambda-C13; vector: plasmid pGSG501; plasmid pGSG502;  
 KM gene transfer; crop improvement; storage carbohydrate; pasture;  
 KM feedstuff; senescence; dextran; binder; food; pharmaceutical.  
 XX  
 OS *Streptococcus salivarius* strain ATCC 25975.

XX W09606173-A1.

XX 29-FEB-1996.

XX 24-AUG-1995; 95WO-AU00527.

XX 24-AUG-1994; 94AU-0007643.

PA (GIF/) GIFFARD P M.  
 XX (JACO/) JACQUES N A.  
 XX (SIMP/) SIMPSON C L.

PI Giffard PM, Jacques NA, Simpson CL;

XX WPI; 1996-151376/15.

DR N-PSDB; AAT13139.

XX  
 PT Plants cont. new bacterial DNA encoding glucosyl transferase  
 PT activity - retain higher levels of stored carbohydrate(s) in a form  
 PT readily digestible by ruminants

XX Claim 4: Page 16-20; 31pp; English.

XX The sequence represents an alpha-D-glucosyltransferase from  
 CC *Streptococcus salivarius*. The enzyme is primer-independent, and  
 CC produces soluble glucan from sucrose. A gene encoding the enzyme  
 CC may be cloned and expressed in *Escherichia coli* using a subclone  
 CC of phage lambda-C13, e.g. plasmid pGSG501 or plasmid pGSG502. The  
 CC DNA may also be expressed in a transgenic plant, to improve the  
 CC level of stored carbohydrate in a pasture plant which normally  
 CC contains low levels, or to prevent degradation of stored carbohydrate  
 CC during plant senescence. Dextran may be isolated from the plant, for  
 CC use as a food binder or pharmaceutical additive. Primer independence  
 CC ensures that the enzyme will be functional in plants. The glucan is

CC poorly degraded in plants but easily degraded by bacteria in the rumen  
 CC of grazing livestock.

XX Sequence 1577 AA:

Query Match 52.9%; Score 63; DB 17; Length 1577;  
 Best Local Similarity 72.2%; Pred. No. 0.74;  
 Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 HLSLEAWSNDPPLYLHD 21  
 |||:|||||  
 Db 591 hlsleawsyndpnykd 608

## RESULT 3

AAB10667  
 ID AAB10667 standard; Protein; 2057 AA.

AC AAB10667;

DT 19-JAN-2001 (first entry)

DE L. mesenteroides alternan sucrose protein.

XX Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;  
 KM syrup.

XX Leuconostoc mesenteroides.

XX DE19905069-A1.

XX 10-AUG-2000.

XX 08-FEB-1999; 99DE-1005069.

XX 08-FEB-1999; 99DE-1005069.

PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.  
 XX (PLAC) MAX PLANCK GES FÖRDERUNG WISSENSCHAFTEN.

XX Kossman J, Welsh T, Quanz M, Knuth K;

XX WPI; 2000-550294/51.

DR N-PSDB; AAA97904.

XX New nucleic acid encoding recombinant Leuconostoc mesenteroides  
 PT alternan sucrose protein and methods of alternan and fructose  
 PT production

XX Claim 1a: Page 30-36; 64pp; German.

XX This invention describes a novel nucleic acid molecule (I) encoding an  
 CC alternan sucrose (E.C. 2.4.1.140 - an enzyme, that belongs to the  
 CC glucosyltransferase group) The recombinant, purified alternan sucrose  
 CC gene is useful for the fermentative production of alternan (a  
 CC carbohydrate) and/or fructose by secreting the enzyme into a  
 CC saccharose-containing culture medium. Alternatively, the enzyme is  
 CC contacted with a saccharose-containing solution. The alternan and/or  
 CC fructose is then isolated from the medium. Cosmetic products or  
 CC foodstuffs containing alternan can be produced. Recombinant production of  
 CC alternan sucrose is advantageous as it provides a cost effective means of  
 CC producing fructose for high fructose containing syrups, production of  
 CC which previously has been achieved by costly production from maize  
 CC starch. This sequence represents the leuconostoc mesenteroides alternan  
 CC sucrose protein which is described in the method of the invention.

XX Sequence 2057 AA:

Query Match 50.4%; Score 60; DB 21; Length 2057;  
 Best Local Similarity 55.0%; Pred. No. 2.8;  
 Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Oy 1 ANDHLSLEAMSDNDTPYLH 20  
||| ||||| | : :  
Db 665 ankhsliedwngkdqyv 684

RESULT 4

AAB36552  
ID AAB36552 standard; Protein: 486 AA.

AC AAB36552;

DT 07-MAR-2001 (first entry)

DE Lawsonia intracellularis flagellar hook protein FlgE SEQ ID NO:1.

KW Lawsonia intracellularis; flgE; flagellar hook protein; vaccine;  
KW intestinal disease; immunogenic; diagnosis; antibacterial; swine;  
KW pig; infection; detection; identification.

XX Lawsonia intracellularis.

WO200069904-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-AU00437.

XX 13-MAY-1999; 99US-0133973.

XX (PF12 ) PFIZER PROD INC.  
PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
PA (PIGR-) PIG RES & DEV CORP.

XX Panaccio M, Rosey EL, Sinistaj M, Hasse D, Parsons J;  
PI Ankenbauer RG;

XX WPI: 2001-016210/02.

DR N-PSDB: AAC88037.

XX New immunogenic Lawsonia FlgE peptide, its nucleic acid and antibody,  
PT useful in vaccines and diagnosis of Lawsonia infections, particularly  
PT in swine -

XX Claim 13; Page 87-90; 97pp; English.

XX The present sequence is the Lawsonia intracellularis flagellar hook  
CC protein FlgE. The present invention describes an isolated or recombinant  
CC polypeptide (I) that comprises, mimics or cross-reacts with a B- or  
CC T-cell epitope of a FlgE (flagellar hook) polypeptide from a

CC Lawsonia spp. (I) has antibacterial activity, and induces a specific  
CC humoral immune response. (I) are used as antigens in vaccines to prevent  
CC or treat infection by Lawsonia, in birds and animals, especially pigs,  
CC to raise specific antibodies (Ab) and to detect past or present  
CC infection. Ab are also useful in diagnosis, to detect L. intracellularis  
CC or immunologically cross-reactive species, also for identification of  
CC epitopes in FlgE. Vectors that contain nucleic acids (II) encoding (I)  
CC are also useful in genetic vaccines, and fragments of (II) are useful  
CC as primers or probes for detecting L. intracellularis or related  
CC microorganisms, in hybridisation or amplification assays.

XX Sequence 486 AA;

SO Query Match 42.9%; Score 51; DB 22; Length 486;  
Best Local Similarity 47.1%; Pred. No. 12;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Oy 1 ANDHLSLEAMSDNDTP 17  
|| : : || : | | |

Db 169 anpyrtalleswngntp 185

RESULT 5

AAG09845  
ID AAG09845 standard; Protein: 305 AA.

XX AAG09845;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 7935.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126254.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 18-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 04-MAY-1999; 99US-0132484.

XX 05-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 07-MAY-1999; 99US-0132487.

XX 11-MAY-1999; 99US-0132863.

XX 14-MAY-1999; 99US-0134256.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 18-MAY-1999; 99US-0134370.

XX 19-MAY-1999; 99US-0134758.

XX 20-MAY-1999; 99US-0134941.

XX 21-MAY-1999; 99US-0135124.

XX 24-MAY-1999; 99US-0135353.

XX 25-MAY-1999; 99US-0135629.

XX 27-MAY-1999; 99US-0136021.

XX 28-MAY-1999; 99US-0136392.

XX 01-JUN-1999; 99US-0136782.

XX 03-JUN-1999; 99US-0137222.

XX 04-JUN-1999; 99US-0137502.

XX 07-JUN-1999; 99US-0137724.

XX 08-JUN-1999; 99US-0138094.

XX 10-JUN-1999; 99US-0138540.

XX 10-JUN-1999; 99US-0138847.

XX 14-JUN-1999; 99US-0139119.

XX 16-JUN-1999; 99US-0139452.

XX 17-JUN-1999; 99US-0139453.

XX 18-JUN-1999; 99US-0139454.

XX 18-JUN-1999; 99US-0139455.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139457.

XX 18-JUN-1999; 99US-0139458.

PR	18-JUN-1999	9905-0139459
PR	18-JUN-1999	9905-0139460
PR	18-JUN-1999	9905-0139461
PR	18-JUN-1999	9905-0139462
PR	18-JUN-1999	9905-0139463
PR	18-JUN-1999	9905-0139465
PR	18-JUN-1999	9905-0139470
PR	21-JUN-1999	9905-0139617
PR	22-JUN-1999	9905-0139699
PR	23-JUN-1999	9905-0140353
PR	24-JUN-1999	9905-0140354
PR	28-JUN-1999	9905-0140695
PR	29-JUN-1999	9905-0140823
PR	30-JUN-1999	9905-0140891
PR	01-JUL-1999	9905-0141287
PR	01-JUL-1999	9905-0141284
PR	02-JUL-1999	9905-0142055
PR	06-JUL-1999	9905-0142390
PR	08-JUL-1999	9905-0142803
PR	09-JUL-1999	9905-0142820
PR	12-JUL-1999	9905-0142977
PR	13-JUL-1999	9905-0143342
PR	14-JUL-1999	9905-0143524
PR	15-JUL-1999	9905-0144005
PR	16-JUL-1999	9905-0144085
PR	16-JUL-1999	9905-0144086
PR	19-JUL-1999	9905-0144325
PR	19-JUL-1999	9905-0144331
PR	19-JUL-1999	9905-0144332
PR	20-JUL-1999	9905-0144332
PR	20-JUL-1999	9905-0144684
PR	21-JUL-1999	9905-0144814
PR	21-JUL-1999	9905-0145088
PR	21-JUL-1999	9905-0145088
PR	22-JUL-1999	9905-0145087
PR	22-JUL-1999	9905-0145087
PR	22-JUL-1999	9905-0145089
PR	22-JUL-1999	9905-0145192
PR	23-JUL-1999	9905-0145145
PR	23-JUL-1999	9905-0145218
PR	23-JUL-1999	9905-0145224
PR	26-JUL-1999	9905-0145576
PR	27-JUL-1999	9905-0145573
PR	27-JUL-1999	9905-0145518
PR	27-JUL-1999	9905-0145519
PR	28-JUL-1999	9905-0145551
PR	02-AUG-1999	9905-0146586
PR	02-AUG-1999	9905-0146588
PR	02-AUG-1999	9905-0146589
PR	03-AUG-1999	9905-0146708
PR	04-AUG-1999	9905-0147204
PR	04-AUG-1999	9905-0147302
PR	05-AUG-1999	9905-0147192
PR	05-AUG-1999	9905-0147260
PR	06-AUG-1999	9905-0147303
PR	06-AUG-1999	9905-0147416
PR	09-AUG-1999	9905-0147493
PR	09-AUG-1999	9905-0147535
PR	10-AUG-1999	9905-0148171
PR	11-AUG-1999	9905-0148191
PR	12-AUG-1999	9905-0148341
PR	13-AUG-1999	9905-0148565
PR	13-AUG-1999	9905-0148684
PR	16-AUG-1999	9905-0149368
PR	17-AUG-1999	9905-0149175
PR	18-AUG-1999	9905-0149426
PR	20-AUG-1999	9905-0149722
PR	20-AUG-1999	9905-0149723

PR	23-AUG-1999	9905-01.499902
PR	23-AUG-1999	9905-01.499922
PR	23-AUG-1999	9905-01.499930
PR	23-AUG-1999	9905-01.500666
PR	26-AUG-1999	9905-01.500884
PR	27-AUG-1999	9905-01.511065
PR	27-AUG-1999	9905-01.511665
PR	27-AUG-1999	9905-01.510800
PR	30-AUG-1999	9905-01.513103
PR	31-ANG-1999	9905-01.514388
PR	01-SEP-1999	9905-01.519330
PR	07-SEP-1999	9905-01.523630
PR	10-SEP-1999	9905-01.530750
PR	13-SEP-1999	9905-01.537588
PR	15-SEP-1999	9905-01.540438
PR	16-SEP-1999	9905-01.540339
PR	20-SEP-1999	9905-01.551739
PR	22-SEP-1999	9905-01.551139
PR	22-SEP-1999	9905-01.534886
PR	24-SEP-1999	9905-01.556559
PR	28-SEP-1999	9905-01.565458
PR	29-SEP-1999	9905-01.565966
PR	04-OCT-1999	9905-01.571177
PR	05-OCT-1999	9905-01.577533
PR	06-OCT-1999	9905-01.578655
PR	07-OCT-1999	9905-01.580829
PR	08-OCT-1999	9905-01.583832
PR	12-OCT-1999	9905-01.585266
PR	13-OCT-1999	9905-01.592934
PR	13-OCT-1999	9905-01.592923
PR	13-OCT-1999	9905-01.595684
PR	14-OCT-1999	9905-01.595329
PR	14-OCT-1999	9905-01.593930
PR	14-OCT-1999	9905-01.596331
PR	14-OCT-1999	9905-01.596337
PR	14-OCT-1999	9905-01.596337
PR	18-OCT-1999	9905-01.595684
PR	21-OCT-1999	9905-01.607441
PR	21-OCT-1999	9905-01.607457
PR	21-OCT-1999	9905-01.607658
PR	21-OCT-1999	9905-01.607770
PR	21-OCT-1999	9905-01.608014
PR	21-OCT-1999	9905-01.608015
PR	22-OCT-1999	9905-01.608850
PR	22-OCT-1999	9905-01.609881
PR	22-OCT-1999	9905-01.609881
PR	22-OCT-1999	9905-01.614004
PR	23-OCT-1999	9905-01.614005
PR	23-OCT-1999	9905-01.614006
PR	26-OCT-1999	9905-01.613559
PR	26-OCT-1999	9905-01.613560
PR	26-OCT-1999	9905-01.613611
PR	26-OCT-1999	9905-01.613920
PR	28-OCT-1999	9905-01.613992
PR	28-OCT-1999	9905-01.621443
PR	29-OCT-1999	9905-01.621993

Query Match	42.0%	Score 50;	DB 21;	Length 305;
Best Local Similarity	52.9%	Pred. No. 10;		
Matches	9;	Conservative	3;	Mismatches
			5;	Indels
				0;
				Gaps
				0;
Oy	3	DHLSTLEAMSDNDPPYL	19	
	:	: : : :	:	
Db	200	dyeqimeawsdkgilyv	216	
RESULT	6			
AAAG09844				
XX	AAAG09844	standard;	Protein;	339 AA.
XX				
AC	AAAG09844;			
XX				
DT	17-OCT-2000	(first entry)		

XX Arabidopsis thaliana protein fragment SEQ ID NO: 7934.  
DE XX  
XX Protein identification: signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
OS Arabidopsis thaliana.  
PN EPI033405-A2.  
XX  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147132.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.

Query Match	42.0%;	Score 50;	DB 21;	Length 339;
Best Local Similarity	52.9%;	Pred. No. 12;		
Matches	9;	Conservative	3;	Mismatches
			5;	Indels
			0;	Gaps
Oy	3	DHLSTLEAMSDNDPPYL 19		
	:		:	
Db	234	dyeimeawsdskgtlyv 250		
RESULT	7			
AAAM43544				
ID	AAAM43544	standard; Protein: 401 AA.		
XX				
AC	AAAM43544;			
XX				
DT	22-Oct-2001	(first entry)		
XX				
DE	Human polypeptide SEQ	ID NO 222.		
XX				
KM	Human: antiarthritic; antirheumatic; antiproliferative; vasotropic;			
KM	cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;			
KM	fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;			
KM	neuroprotective; antiallergic; hepatotropic; antidiabetic;			

KW		antiinflammatory; antiulcer; vulnereary; anticonvulsant; antibacterial,
KV		antiparasitic; cardiant; gene therapy; cancer; immune disorder;
KW		cardiovascular disorder; neurological disease; infection; human.
XX		
OS	Homo sapiens.	
XX		
PN	WO200155308-A2.	
XX		
PD	02-AUG-2001.	
XX		
PF	17-JAN-2001; 2001WO-US01309.	
XX		
PR	31-JAN-2000; 2000US-0179065.	
PR	04-FEB-2000; 2000US-0180628.	
PR	24-FEB-2000; 2000US-0184664.	
PR	02-MAR-2000; 2000US-0186350.	
PR	16-MAR-2000; 2000US-0189874.	
PR	17-MAR-2000; 2000US-0190076.	
PR	18-APR-2000; 2000US-0198123.	
PR	19-MAY-2000; 2000US-0205515.	
PR	07-JUN-2000; 2000US-0209467.	
PR	28-JUN-2000; 2000US-0214886.	
PR	30-JUN-2000; 2000US-0215135.	
PR	07-JUL-2000; 2000US-0216647.	
PR	07-JUL-2000; 2000US-0216880.	
PR	11-JUL-2000; 2000US-0217487.	
PR	11-JUL-2000; 2000US-0217496.	
PR	14-JUL-2000; 2000US-0218290.	
PR	26-JUL-2000; 2000US-0220963.	
PR	26-JUL-2000; 2000US-0220964.	
PR	14-AUG-2000; 2000US-0224518.	
PR	14-AUG-2000; 2000US-0224519.	
PR	14-AUG-2000; 2000US-0225213.	
PR	14-AUG-2000; 2000US-0225214.	
PR	14-AUG-2000; 2000US-0225266.	
PR	14-AUG-2000; 2000US-0225267.	
PR	14-AUG-2000; 2000US-0225268.	
PR	14-AUG-2000; 2000US-0225270.	
PR	14-AUG-2000; 2000US-0225447.	
PR	14-AUG-2000; 2000US-0225757.	
PR	14-AUG-2000; 2000US-0225758.	
PR	14-AUG-2000; 2000US-0225759.	
PR	18-AUG-2000; 2000US-0226279.	
PR	22-AUG-2000; 2000US-0226681.	
PR	22-AUG-2000; 2000US-0226688.	
PR	22-AUG-2000; 2000US-0227182.	
PR	23-AUG-2000; 2000US-0227009.	
PR	30-AUG-2000; 2000US-0228924.	
PR	01-SEP-2000; 2000US-0229287.	
PR	01-SEP-2000; 2000US-0229343.	
PR	01-SEP-2000; 2000US-0229344.	
PR	01-SEP-2000; 2000US-0229345.	
PR	05-SEP-2000; 2000US-0229509.	
PR	05-SEP-2000; 2000US-0229513.	
PR	06-SEP-2000; 2000US-0230437.	
PR	06-SEP-2000; 2000US-0230438.	
PR	08-SEP-2000; 2000US-0231242.	
PR	08-SEP-2000; 2000US-0231243.	
PR	08-SEP-2000; 2000US-0231244.	
PR	08-SEP-2000; 2000US-0231413.	
PR	08-SEP-2000; 2000US-0231414.	
PR	08-SEP-2000; 2000US-0232080.	
PR	08-SEP-2000; 2000US-0232081.	
PR	12-SEP-2000; 2000US-0231968.	
PR	14-SEP-2000; 2000US-0232397.	
PR	14-SEP-2000; 2000US-0232398.	
PR	14-SEP-2000; 2000US-0232399.	
PR	14-SEP-2000; 2000US-0232400.	
PR	14-SEP-2000; 2000US-0232401.	
PR	14-SEP-2000; 2000US-0233063.	
PR	14-SEP-2000; 2000US-0233064.	
PR	14-SEP-2000; 2000US-0233065.	
PR	21-SEP-2000; 2000US-0234223.	



[illegible]

```

PR 11-DEC-2000; 20000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI: 2001-488781/53.
XX
DR N-PSDB; AA163850.
XX
PT New isolated nucleic acids and polypeptides, useful for diagnosing,
PT treating and/or preventing human diseases and disorders -
XX
PS Claim 11: SEQ ID NO 222; 664bp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AA163803-AA164012) and
CC the encoded proteins (AAM34497-AAM3660) useful for preventing, treating
CC or ameliorating medical conditions e.g. by protein or gene therapy. The
CC genes were isolated from a range of human tissues disclosed in the
CC specification. The nucleic acids, proteins, antibodies and (ant)agonists
CC are useful in the diagnosis, treatment and prevention of: (a) cancer,
CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,
CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,
CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 401 AA:
XX
Query Match 40.8%; Score 48.5; DB 22; Length 401;
Best Local Similarity 47.4%; Pred. No. 24;
Matches 9; Conservative 3; Mismatches 4; Indels 3; Gaps 1;
OY 5 ISLEAWSNDN--DPPYLH 20
:::| | | | |
Db 46 vallevwkkdnrfddxpxlh 64
RESULT 8
AAV67414
ID AAV67414 standard; protein: 534 AA.
XX
XX
AC AAV67414;
XX
DT 12-MAY-2000 (first entry)
XX
DE Arabidopsis aldehyde dehydrogenase (ALDH)-3.
XX
XX Arabidopsis: plant plastid; acetyl CoA synthetase; E3 subunit; ACS;
KW plant plastidic pyruvate dehydrogenase; PPDH; ATP citrate lyase; ACL;
KW pyruvate decarboxylase; PDC; aldehyde dehydrogenase; ALDH; acetyl CoA;
KW fatty acid; flavonoid; enzyme; phytochemical; pyruvate decarboxylase;
KW acetyl CoA hydrolase; mitochondrial pyruvate dehydrogenase.
XX
OS Arabidopsis sp.
XX
PN WO200000619-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14382.
XX
PR 26-JUN-1998; 98US-0090717.
XX
PA (IOWA) UNIV IOWA STATE RES FOUND INC.

```

XX Nikolau BJ, Murtelle ES, Oliver DJ, Behal R, Schnable PS, Ke J;  
PI Johnson JL, Allred CC, Fatland B, Lutziger I, Wen T;  
XX  
DR WPI: 2000-160678/14.  
N-PSDB: AA556975.  
XX  
XX Novel acetyl CoA synthetase (ACS), plastidic pyruvate dehydrogenase  
PT (PPDH), Arp citrate lyase (ACL), pyruvate decarboxylase (PDC) and  
PT aldehyde dehydrogenase (ALDH) polypeptides used to alter acetyl CoA  
PT levels in plants  
XX  
XX Examples: Fig 12B; 79pp; English.  
XX  
XX The invention provides nucleic acids encoding Arabidopsis plastidic  
CC acetyl CoA synthetase (ACS), various subunits (specifically the E3  
CC subunit) of plant plastidic pyruvate dehydrogenase (PPDH), the A and B  
CC subunits of a plant Arp citrate lyase (ACL), Arabidopsis pyruvate  
CC decarboxylase (PDC), Arabidopsis aldehyde dehydrogenase (ALDH),  
CC specifically ALDH-2 and ALDH-4. The polypeptides can be expressed by  
CC polypeptides, methods and nucleic acid molecules of the invention are  
CC used to alter the level of acetyl CoA in a plant or plant cell, tissues  
CC or organs. A decrease in acetyl CoA is expected to affect the  
CC biosynthesis of very long chain fatty acids and flavonoids. The enzymes  
CC may also be used for in vitro synthesis of acetyl CoA, which in turn can  
CC be used to produce acetyl CoA phytochemicals (plastidic ACS, PPDH, ACL,  
CC pyruvate decarboxylase, acetyl CoA hydrolyase, mitochondrial pyruvate  
CC dehydrogenase and aldehyde dehydrogenase are involved in the biosynthesis  
CC of acetyl CoA.  
XX  
SQ Sequence 534 AA:

Query Match 39.9%; Score 47.5; DB 21; Length 534;  
Best Local Similarity 52.9%; Pred. No. 47;  
Matches 9; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 2 NDHLSTLEAMSDNTPY 18  
||::|||  
DB 132 ndelaaletw-dngkpy 147

RESULT 9  
AAG09941  
ID AAG09941 strand: Protein; 195 AA.  
AAG09941;  
XX  
XX 17-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 8067.  
DE  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
XX Arabidopsis thaliana.  
OS  
XX  
XX EPI033405-A2.  
PN  
XX  
XX 06-SEP-2000.  
PD  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PF  
XX  
XX 25-FEB-1999; 990S-0121825.  
PR 05-MAR-1999; 990S-0123180.  
PR 09-MAR-1999; 990S-0123548.  
PR 23-MAR-1999; 990S-0125788.  
PR 25-MAR-1999; 990S-0126264.  
PR 29-MAR-1999; 990S-0126785.  
PR 01-APR-1999; 990S-0127462.  
PR 06-APR-1999; 990S-0128234.

PR 08-APR-1999; 990S-0128714.  
PR 16-APR-1999; 990S-0129845.  
PR 19-APR-1999; 990S-0130077.  
PR 21-APR-1999; 990S-0130449.  
PR 23-APR-1999; 990S-0130510.  
PR 28-APR-1999; 990S-0130891.  
PR 30-APR-1999; 990S-0131449.  
PR 30-APR-1999; 990S-0132048.  
PR 04-MAY-1999; 990S-0132407.  
PR 05-MAY-1999; 990S-0132484.  
PR 06-MAY-1999; 990S-0132486.  
PR 06-MAY-1999; 990S-0132487.  
PR 07-MAY-1999; 990S-0132863.  
PR 11-MAY-1999; 990S-0134256.  
PR 14-MAY-1999; 990S-0134218.  
PR 14-MAY-1999; 990S-0134219.  
PR 14-MAY-1999; 990S-0134221.  
PR 14-MAY-1999; 990S-0134370.  
PR 18-MAY-1999; 990S-0134768.  
PR 19-MAY-1999; 990S-0134941.  
PR 20-MAY-1999; 990S-0135124.  
PR 21-MAY-1999; 990S-0135353.  
PR 24-MAY-1999; 990S-0135629.  
PR 25-MAY-1999; 990S-0136021.  
PR 27-MAY-1999; 990S-0136392.  
PR 28-MAY-1999; 990S-0136782.  
PR 01-JUN-1999; 990S-0137222.  
PR 03-JUN-1999; 990S-0137528.  
PR 04-JUN-1999; 990S-0137502.  
PR 07-JUN-1999; 990S-0137724.  
PR 08-JUN-1999; 990S-0138094.  
PR 10-JUN-1999; 990S-0138540.  
PR 10-JUN-1999; 990S-0138847.  
PR 14-JUN-1999; 990S-0138119.  
PR 16-JUN-1999; 990S-0139452.  
PR 16-JUN-1999; 990S-0139453.  
PR 17-JUN-1999; 990S-0139492.  
PR 18-JUN-1999; 990S-0139454.  
PR 18-JUN-1999; 990S-0139455.  
PR 18-JUN-1999; 990S-0139456.  
PR 18-JUN-1999; 990S-0139457.  
PR 18-JUN-1999; 990S-0139457.  
PR 18-JUN-1999; 990S-0139458.  
PR 18-JUN-1999; 990S-0139459.  
PR 18-JUN-1999; 990S-0139460.  
PR 18-JUN-1999; 990S-0139461.  
PR 18-JUN-1999; 990S-0139462.  
PR 18-JUN-1999; 990S-0139463.  
PR 18-JUN-1999; 990S-0139750.  
PR 18-JUN-1999; 990S-0139763.  
PR 21-JUN-1999; 990S-0139817.  
PR 22-JUN-1999; 990S-0139899.  
PR 23-JUN-1999; 990S-0140353.  
PR 23-JUN-1999; 990S-0140354.  
PR 24-JUN-1999; 990S-0140695.  
PR 28-JUN-1999; 990S-0140823.  
PR 29-JUN-1999; 990S-0140991.  
PR 30-JUN-1999; 990S-0141287.  
PR 01-JUL-1999; 990S-0141842.  
PR 01-JUL-1999; 990S-0142154.  
PR 02-JUL-1999; 990S-0142055.  
PR 06-JUL-1999; 990S-0142390.  
PR 08-JUL-1999; 990S-0142803.  
PR 09-JUL-1999; 990S-0142920.  
PR 12-JUL-1999; 990S-0142977.  
PR 13-JUL-1999; 990S-0143542.  
PR 14-JUL-1999; 990S-0143624.  
PR 15-JUL-1999; 990S-0144005.  
PR 16-JUL-1999; 990S-0144085.  
PR 16-JUL-1999; 990S-0144086.  
PR 19-JUL-1999; 990S-0144325.  
PR 19-JUL-1999; 990S-0144331.  
PR 19-JUL-1999; 990S-0144332.

PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.

PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 39.5%; Score 47; DB 21; Length 195;  
Best Local Similarity 36.8%; Pred. No. 17;  
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 3 DRLSLEAMSDMDTPYLHD 21  
Db 109 nhqevdawshqkplwtd 127

RESULT 10

ID AAR65965 standard; Protein, 338 AA.

AC AAR65965;

DT 06-JUN-1995 (first entry)

XX T. niveum GAPDH.

XX Alanine-racemase; glyceraldehyde-3-phosphate-dehydrogenase;

KW GAPDH; cyclosporin; immunosuppressive.

XX Tolypocladium niveum ATCC 34921.

XX W09425606-A.

PN 10-NOV-1994.

XX PD 23-APR-1994; 94WO-EP01272.

XX PP 23-APR-1993; 93DE-4312856.

XX PR 04-MAY-1993; 93DE-4314610.

XX PR 17-MAY-1993; 93DE-4316419.

XX PA (SANO ) SANDOZ LTD.

XX PA (SANO ) SANDOZ PATENT GMBH.

XX PA (SANO ) SANDOZ-ERFINDUNGEN VERW GRS MBH.

XX PI Kocher HP, Schneider-scherzer E, Schoergendorfer K;

XX PI Weber G;

XX DR WPI: 1994-358281/44.

XX New nucleic acid encoding eukaryotic alanine racemase - and  
PT related vectors, host cells and recombinant enzyme, useful for  
PT producing cyclosporin derivs. or increasing cyclosporin prodn.,





```

XX      Fusarium oxysporum.
OS
XX
FH      Key                      Location/Qualifiers
FT      Domain                  22..54
FT      /label= CBD
FT      /note= "homologous to a terminal A region of
FT      T.reesei cellulase"
XX
PN      WO9117244-A.
XX
PD      14-NOV-1991.
XX
PF      08-MAY-1991; 91WO-DK00124.
XX
PR      09-MAY-1990; 90DK-0001158.
XX
XX      (NOVO ) NOVO NORDISK A/S.
XX
XX      Woldike HF, Hagen F, Hjort CW, Hastrup S;
DR      MPI: 1991-353766/48.
DR      N-ESDB; AAQ14839.
XX
XX      New fungal (hemi) cellulose degrading enzymes - for prodn. of
PT      liq. fuel gas and feed protein, have specified carbohydrate
PT      binding domain
XX
PS      Claim 1; Fig 12; 73pp; English.
XX
CC      The full-length F-family cellobiohydrolase coding sequence was
CC      isolated from a F.oxysporum cDNA library by hybridisation to PCR
CC      generated genomic oligonucleotide probes (see esp. AAQ15504). The
CC      enzyme encoded by the sequence has a CBD which is homologous to a
CC      terminal A region of T.reesei cellulase and is able to bind to
CC      insoluble (hemi)cellulosic substrates. See AAQ14838-Q14842.
XX
SQ      Sequence 385 AA:

```

```

Query Match          37.8%; Score 45; DB 12; Length 385;
Best Local Similarity 24.4%; Pred. No. 77;
Matches 10; Conservative 5; Mismatches 6; Indels 20; Gaps 1;

QY      1 ANDHLSILEA-----WSDNDPPLHD 21
        |||:|:::| | | | | | | |
        329 andyiltwmaclavpkcvgiltwgvskdswrpgdnpllyd 369

```

Search completed: March 27, 2002, 13:57:55  
 Job time: 523 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein : protein search, using sw model

Run on: March 27, 2002, 13:59:28 ; Search time 87.3 Seconds

(without alignments)  
5.413 Million cell updates/sec

Title: US-09-290-049A-10

Sequence: 1 ANDHLSILEAWSNDNTPYLHD 21

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents-AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	1475	US-09-007-999-2	Sequence 2, Appli
2	119	100.0	1475	US-09-210-361-2	Sequence 2, Appli
3	110	92.4	1375	US-09-210-361-4	Sequence 4, Appli
4	75	63.0	1430	US-09-008-172-2	Sequence 2, Appli
5	75	63.0	1430	US-09-210-361-6	Sequence 6, Appli
6	63	52.9	1577	US-08-793-824-2	Sequence 2, Appli
7	45	37.8	385	US-08-361-920-23	Sequence 23, Appli
8	45	37.8	385	US-08-479-939-23	Sequence 23, Appli
9	45	37.8	385	US-08-483-432-23	Sequence 23, Appli
10	44	37.0	384	US-09-311-170-2	Sequence 2, Appli
11	43	36.1	2465	US-08-596-291-3	Sequence 3, Appli
12	43	36.1	2465	US-09-100-804-3	Sequence 3, Appli
13	43	36.1	2466	US-09-080-855-12	Sequence 12, Appli
14	43	36.1	2466	PCT-US94-09943-2	Sequence 2, Appli
15	43	36.1	2485	US-09-290-640-46	Sequence 46, Appli
16	42	35.3	259	US-08-277-231A-3	Sequence 3, Appli
17	42	35.3	259	US-08-473-750-6	Sequence 6, Appli
18	42	35.3	259	US-08-477-326-6	Sequence 6, Appli
19	42	35.3	259	US-08-070-301-16	Sequence 16, Appli
20	41	34.5	480	US-08-272-255-12	Sequence 12, Appli
21	41	34.5	480	PCT-US95-08565-12	Sequence 12, Appli
22	40	33.6	196	US-08-933-750C-35	Sequence 35, Appli
23	40	33.6	196	US-09-234-613-35	Sequence 35, Appli
24	40	33.6	307	US-08-332-562A-83	Sequence 83, Appli
25	40	33.6	334	US-09-362-473-6	Sequence 6, Appli
26	40	33.6	455	US-09-362-473-6	Sequence 6, Appli
27	40	33.6	761	US-09-625-188-14	Sequence 14, Appli

28	40	33.6	839	4	US-09-197-636-2	Sequence 2, Appli
29	40	33.6	839	4	US-09-197-636-4	Sequence 4, Appli
30	40	33.6	839	4	US-09-197-636-8	Sequence 8, Appli
31	40	33.6	2860	2	US-08-826-267-2	Sequence 2, Appli
32	40	33.6	3898	4	US-08-750-717-2	Sequence 2, Appli
33	39.5	33.2	337	3	US-09-032-372-2	Sequence 3, Appli
34	39.5	33.2	500	6	5171684-2	Patent No. 5171684
35	39.5	33.2	567	1	US-08-258-261B-4	Sequence 4, Appli
36	39.5	33.2	567	1	US-08-456-837-4	Sequence 4, Appli
37	39.5	33.2	567	1	US-08-457-342-4	Sequence 4, Appli
38	39.5	33.2	567	1	US-08-457-646A-4	Sequence 4, Appli
39	39.5	33.2	567	1	US-08-458-076A-4	Sequence 4, Appli
40	39.5	33.2	567	1	US-08-457-335A-4	Sequence 4, Appli
41	39.5	33.2	567	2	US-08-729-214-4	Sequence 4, Appli
42	39.5	33.2	567	2	US-09-028-934-4	Sequence 4, Appli
43	39	32.8	755	3	US-09-071-101-2	Sequence 2, Appli
44	39	32.8	755	3	US-09-369-618-2	Sequence 2, Appli
45	39	32.8	755	3	US-09-369-617-2	Sequence 2, Appli

#### ALIGNMENTS

```

RESULT 1
US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0356D
; CURRENT APPLICATION NUMBER: US/09/007,999
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-007-999-2

Query Match 100.0%; Score 119; DB 3; Length 1475;
Best Local Similarity 100.0%; Pred. No. 6.4e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSILEAWSNDNTPYLHD 21
Db 481 ANDHLSILEAWSNDNTPYLHD 501

RESULT 2
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07

```

```
;; EARLIER APPLICATION NUMBER: 09/008,172
;; EARLIER FILING DATE: 1998-01-16
;; EARLIER APPLICATION NUMBER: 08/482,711
;; EARLIER FILING DATE: 1995-06-07
;; NUMBER OF SEQ ID NOS: 6
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 2
;; LENGTH: 1475
;; TYPE: PRF
;; ORGANISM: Streptococcus mutans
US-09-210-361-2
```

```
Query Match          100.0%; Score 119; DB 4; Length 1475;
Best Local Similarity 100.0%; Pred. No. 6,4e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 ANDHLSLEAWSNDPTPLHD 21
    |||
DB 481 ANDHLSLEAWSNDPTPLHD 501
```

```
RESULT 3
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRF
; ORGANISM: streptococcus mutans
US-09-210-361-4
```

```
Query Match          92.4%; Score 110; DB 4; Length 1375;
Best Local Similarity 95.2%; Pred. No. 1.7e-09;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 ANDHLSLEAWSNDPTPLHD 21
    |||
DB 507 ANDHLSLEAWSNDPTPLHD 527
```

```
RESULT 4
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
```

```
;; CURRENT FILING DATE: 1998-01-16
;; EARLIER APPLICATION NUMBER: 08/482,711
;; EARLIER FILING DATE: 1995-06-07
;; NUMBER OF SEQ ID NOS: 2
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 2
;; LENGTH: 1430
;; TYPE: PRF
;; ORGANISM: Streptococcus mutans
US-09-008-172-2
```

```
Query Match          63.0%; Score 75; DB 3; Length 1430;
Best Local Similarity 71.4%; Pred. No. 0.00084;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 1 ANDHLSLEAWSNDPTPLHD 21
    |||
DB 495 AINHLSLEAWSNDPQYKND 515
```

```
RESULT 5
US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRF
; ORGANISM: streptococcus mutans
US-09-210-361-6
```

```
Query Match          63.0%; Score 75; DB 4; Length 1430;
Best Local Similarity 71.4%; Pred. No. 0.00084;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 1 ANDHLSLEAWSNDPTPLHD 21
    |||
DB 495 AINHLSLEAWSNDPQYKND 515
```

```
RESULT 6
US-08-793-824-2
; Sequence 2, Application US/08793824
; Patent No. 5981838
; GENERAL INFORMATION:
; APPLICANT: Simpson, Christine Lynn
; APPLICANT: Giffard, Philip Morrison
; APPLICANT: Jacques, Nicholas Anthony
; TITLE OF INVENTION: Genetic Manipulation of Plants to
; INCREASE STORED CARBOHYDRATES
; NUMBER OF SEQUENCES: 2
```



;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Griffith Hack & Co  
;; STREET: Level 8, 168 Walker Street  
;; CITY: No. 5981838th Sydney  
;; STATE: New South Wales  
;; COUNTRY: Australia  
;; ZIP: 2060  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/793,824  
;; FILING DATE:  
;; CLASSIFICATION: 800  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: AU PM7643  
;; FILING DATE: 24-AUG-1994  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 61 2 9957 5944  
;; TELEFAX: 61 2 957 6288  
;; TELEX: 26547  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1577 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: not relevant  
;; MOLECULE TYPE: protein  
;; ORIGINAL SOURCE:  
;; ORGANISM: Streptococcus salivarius  
;; US-08-793-824-2

Query Match 52.9%; Score 63; DB 2; Length 1577;  
Best Local Similarity 72.2%; Pred. No. 0.084;  
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 HSLILEMSNDPPLYLD 21  
||||| | | | |  
Db 591 HSLILEMSYNDHQYKND 608

RESULT 7  
US-08-361-920-23  
; Sequence 23, Application US/08361920  
; Patent No. 5457046  
; GENERAL INFORMATION:  
; APPLICANT: Woeldike, Helle F.  
; APPLICANT: Hagen, Frederick  
; APPLICANT: Hjort, Carsten M.  
; APPLICANT: Sven, Hastrup  
; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose  
; NUMBER OF SEQUENCES: 85  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 54570460 No. 5457046disk of No. 5457046th America, Inc.  
; STREET: 405 Lexington Avenue, 62nd Floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 10174-6201  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/361,920  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/940,860  
;; FILING DATE: 28-OCT-1992  
;; APPLICATION NUMBER: DK 1158/90  
;; FILING DATE: 09-MAY-1990  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/DK91/00124  
;; FILING DATE: 08-MAY-1991  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Lambiris, Elias J.  
;; REGISTRATION NUMBER: 33,728  
;; REFERENCE/DOCKET NUMBER: 3435, 204-US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 212-867-0123  
;; TELEFAX: 212-867-0298  
;; INFORMATION FOR SEQ ID NO: 23:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 385 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-361-920-23

Query Match 37.8%; Score 45; DB 1; Length 385;  
Best Local Similarity 24.4%; Pred. No. 13;  
Matches 10; Conservative 5; Mismatches 6; Indels 20; Gaps 1;

OY 1 ANDHLSLEA-----MSNDPPLYLD 21  
|||:|:| | | | |  
Db 329 ANDYLTVMNACLAVPKCVGITWVGVSDBKDSWPGDNPPLYLD 369

RESULT 8  
US-08-479-939-23  
; Sequence 23, Application US/08479939  
; Patent No. 5686593  
; GENERAL INFORMATION:  
; APPLICANT: Woeldike, Helle F.  
; APPLICANT: Hagen, Frederick  
; APPLICANT: Hjort, Carsten M.  
; APPLICANT: Sven, Hastrup  
; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose  
; NUMBER OF SEQUENCES: 85  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 56865930 No. 5686593disk of No. 5686593th America, Inc.  
; STREET: 405 Lexington Avenue, 62nd Floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 10174-6201  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/479,939  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/361,920  
; FILING DATE: 22-DEC-1994  
; APPLICATION NUMBER: US 07/940,860  
; FILING DATE: 28-OCT-1992  
; APPLICATION NUMBER: DK 1158/90  
; FILING DATE: 09-MAY-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/DK91/00124  
; FILING DATE: 08-MAY-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lambiris, Elias J.  
; REGISTRATION NUMBER: 33,728

```

? REFERENCE/DOCKET NUMBER: 3435..204-US
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 212-867-0123
? TELEFAX: 212-867-0298
? INFORMATION FOR SEO ID NO: 23:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 385 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-479-939-23

```

Query Match	37.8%	Score 45	DB 1	Length 365
Best Local Similarity	24.4%	Pred. No. 13		
Matches 10	Conservative 5	Mismatches 6	Indels 20	Gaps 1

```

1  ANDHLSILEA-----WSDNDTPYLHD 21
   |||::: | | | |
329 ANDYLTVMNACLAIVKCVGITVWGVSDKDSWRPGDNPLLYD 365

```

RESULT 9  
US-08-483-432-23  
; Sequence 23, Application US/08483432  
; Patent No. 5763254

1  
 2 GENERAL INFORMATION:  
 3 APPLICANT: Woeldike, Helle F.  
 4 APPLICANT: Hagen, Frederick  
 5 APPLICANT: Hjort, Carsten M.  
 6 APPLICANT: Sven, Hastrup  
 7 TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose  
 8 TITLE OF INVENTION: Or Hemicellulose  
 9 NUMBER OF SEQUENCES: 85  
 10 CORRESPONDENCE ADDRESS:  
 11 ADDRESSEE: NO. 57632540 NO. 5763254disk of NO. 5763254th America, Inc

```

1 ZIP: 10174-6201
2
3 COMPUTER READABLE FORM:
4
5 MEDIUM TYPE: Floppy disk
6
7 COMPUTER: IBM PC compatible
8
9 OPERATING SYSTEM: PC-DOS/MS-DOS
10
11 SOFTWARE: PatentIn Release #1.0, Version #1.2
12
13 CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/483,432  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/361,920  
FILING DATE:  
APPLICATION NUMBER: US 07/940,860  
FILING DATE: 28-OCT-1992  
APPLICATION NUMBER: DK 1158/90  
FILING DATE: 09-MAY-1990

1 PRIOR APPLICATION DATA:  
 2 APPLICATION NUMBER: PCT/DK91/00124  
 3 FILING DATE: 08-MAY-1991  
 4 ATTORNEY/AGENT INFORMATION:  
 5 NAME: Lambiris, Elias J.  
 6 REGISTRATION NUMBER: 33,728  
 7 REFERENCE/DOCKET NUMBER: 345.204-US  
 8 TELECOMMUNICATION INFORMATION:  
 9 TELEPHONE: 212-867-0123  
 0 TELEFAX: 212-867-0298  
 1 INFORMATION FOR SEQ. ID NO.: 23:  
 2 (SEQUENCE CHARACTERISTICS)

```

;
; SEQUENCE CHARACTERISTICS
; LENGTH: 385 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

US-08-483-432-23

Query Match	37.88;	Score 45;	DB 1;	Length 385;
Best Local Similarity	24.48;	Pred. No. 13;		
Matches 10;	Conservative 5;	Mismatches 6;	Indels 20;	Gaps 1;

```

QY      1 ANDHSLLEA-----WSDNDTPYLHD 21
        |||:::|      | | | |
Db      329 ANDYLTVMNACLAVPKCVGJITVGVSDKDSWRPGDNPPLVD 365

```

RESULT 10  
US-09-311-170-2

```

: sequence 2, application US/09311170
: Patent No. 6121034
: GENERAL INFORMATION:
: APPLICANT: Laroche et al., Andre L.
: TITLE OF INVENTION: xylanase cxyl1
: FILE REFERENCE: xylanase cxyl1
: CURRENT APPLICATION NUMBER: US/09/311,170
: CURRENT FILING DATE: 1999-05-13
: NUMBER OF SEQ ID NOS: 7
: SOFTWARE: PatentIn Ver. 2.0

```

```

; TYPE: PRT
; ORGANISM: Coniothyrium minitans
US-09-311-170-2

```

Query Match	37.0%;	Score 44;	DB 3;	Length 384;
Best Local Similarity	22.0%;	Pred. No. 19;		
Matches	9;	Conservative	8;	Mismatches 4;
				Indels 20;
				Gaps 1;

```
QY      1 ANDHLSIL-----EAWSDNDTPYLHD 21
      |||:::      ::|||:|
Db      328 ANDYLTVMNGCLAVPKCVGITVMGVSDDKSDSRSSDPLLED 368
```

RESULT 11  
US-08-596-291-3  
; Sequence 3, Application US/08596291  
; Patent No. 5821075

```

1  GENERAL INFORMATION:      LEONEL JORGE
2  APPLICANT: GONEZ,
3  APPLICANT: SARAS, JAN
4  APPLICANT: CLAESSON-WELSH, LENA
5  APPLICANT: HELDIN, CARL-HENRIK
6  TITLE OF INVENTION:      PRIMARY STRUCTURE AND FUNCTIONAL
7  TITLE OF INVENTION:      EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
8  TITLE OF INVENTION:      TYROSINE PHOSPHATASES
9  NUMBER OF SEQUENCES:      4

```

1 CORRESPONDENCE ADDRESS:  
2 ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
3 STREET: 600 ATLANTIC AVENUE  
4 CITY: BOSTON  
5 STATE: MASSACHUSETTS  
6 COUNTRY: USA

```

1 ZIP: 02210
2
3 COMPUTER READABLE FORM:
4
5 MEDIUM TYPE: Floppy disk
6
7 COMPUTER: IBM PC compatible
8
9 OPERATING SYSTEM: PC-DOS/MS-DOS
10
11 SOFTWARE: PatentIn Release #1.0, Version #1.25
12
13 CURRENT APPLICATION DATA:
14
15 APPLICATION NUMBER: US/08/596,291
16
17 FILING DATE: 09-AUG-1996

```

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/115,573  
FILING DATE: 01-SEP-1993  
ATTORNEY/AGENT INFORMATION:



CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/09943  
FILING DATE: 01-SEP-1994  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/115,573  
FILING DATE: 01-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: TWOMEY, MICHAEL J.  
REGISTRATION NUMBER: P-38,349  
REFERENCE/DOCKET NUMBER: LC461/700000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/720-3500  
TELEFAX: 617/720-2441  
TELEX: 92-1742 EZEKIEL  
INFORMATION FOR SEQ. ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2466 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-09943-2

Query Match 36.1%; Score 43; DB 5; Length 2466;  
Best Local Similarity 56.2%; Pred. No. 2.5e+02;  
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

OY 4 HLSILE--AMSDNDTP 17  
|:|:| |1:1:11|  
DB 2348 HISHLNFMTAMPDHDTP 2363

RESULT 15  
US-09-290-640-46  
Sequence 46, Application US/09290640  
Patent No. 6204055  
GENERAL INFORMATION:  
APPLICANT: Dean, Nicholas M.  
TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signaling  
FILE REFERENCE: ISPH-0351  
CURRENT APPLICATION NUMBER: US/09/290,640  
CURRENT FILING DATE: 1999-04-12  
NUMBER OF SEQ. ID NOS: 85  
SOFTWARE: PatentIn Ver. 2.0  
SEQ. ID NO 46  
LENGTH: 2485  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-290-640-46

Query Match 36.1%; Score 43; DB 4; Length 2485;  
Best Local Similarity 56.2%; Pred. No. 2.5e+02;  
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;  
OY 4 HLSILE--AMSDNDTP 17  
|:|:| |1:1:11|  
DB 2367 HISHLNFMTAMPDHDTP 2382

Search completed: March 27, 2002, 13:59:29  
Job time: 582 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:20:24 ; Search time 1139.61 seconds

(without alignments)  
5.116 Million cell updates/sec

Title: US-09-290-049a-10

Sequence: 1 ANDHLSTLEAWSNDPTPLD 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

1 number of hits satisfying chosen parameters: 3148936

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/2/paa/PCTRUS\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US080\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep:\*  
8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep:\*  
9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep:\*  
10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep:\*  
11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep:\*  
12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep:\*  
13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep:\*  
14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep:\*  
15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep:\*  
16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep:\*  
17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep:\*  
18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep:\*  
19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep:\*  
20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep:\*  
21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep:\*  
22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep:\*  
23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep:\*  
24: /cgn2\_6/ptodata/2/paa/US60\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	21	US-09-290-049-1	Sequence 1, Appl1
2	119	100.0	21	US-09-290-049-10	Sequence 10, Appl1
3	119	100.0	1475	US-09-557-848-2	Sequence 2, Appl1
4	119	100.0	1475	US-09-740-274-2	Sequence 2, Appl1
5	110	92.4	1375	US-09-740-274-4	Sequence 4, Appl1
6	108	90.8	21	US-09-290-049-12	Sequence 12, Appl1
7	108	90.8	21	US-09-290-049-14	Sequence 14, Appl1
8	75	63.0	21	US-09-290-049-11	Sequence 11, Appl1
9	75	63.0	1430	US-09-649-885-2	Sequence 2, Appl1

10	75	63.0	1430	21	US-09-740-274-6	Sequence 6, Appl1
11	73	61.3	21	16	US-09-290-049-13	Sequence 13, Appl1
12	60	50.4	2057	18	US-09-499-203-2	Sequence 2, Appl1
13	50	42.0	306	19	US-09-595-298A-661	Sequence 661, Appl1
14	50	42.0	340	19	US-09-595-298A-660	Sequence 660, Appl1
15	49.5	41.6	546	24	US-60-324-109-17015	Sequence 17015, A
16	49.5	41.6	549	24	US-60-324-109-16789	Sequence 16789, A
17	49	41.2	391	24	US-60-324-109-20736	Sequence 20736, A
18	48.5	40.8	401	1	PCT-US01-01309-222	Sequence 222, App
19	48	40.3	773	16	US-09-248-796-22573	Sequence 22573, A
20	47.5	39.9	178	22	US-09-867-716-19848	Sequence 19848, A
21	47.5	39.9	534	14	US-09-344-882-24	Sequence 24, Appl1
22	47.5	39.9	534	24	US-60-324-109-33149	Sequence 33149, A
23	47	39.5	196	19	US-09-595-298A-765	Sequence 765, App
24	46.5	39.1	220	24	US-60-212-413-164	Sequence 164, App
25	46.5	39.1	220	24	US-60-229-518-331	Sequence 331, App
26	46.5	39.1	330	24	US-60-207-422-75	Sequence 75, Appl1
27	46.5	39.1	539	1	PCT-US01-01332-762	Sequence 762, App
28	46.5	39.1	539	21	US-09-764-875-762	Sequence 762, App
29	46	38.7	62	18	US-09-417-507-26597	Sequence 26597, A
30	46	38.7	94	1	PCT-US01-08656-5595	Sequence 5595, Ap
31	46	38.7	118	20	US-09-673-395A-221	Sequence 221, App
32	46	38.7	199	21	US-09-738-626-4571	Sequence 4571, Ap
33	46	38.7	275	19	US-09-538-092-348	Sequence 348, App
34	46	38.7	336	15	US-09-198-452A-662	Sequence 662, App
35	46	38.7	336	18	US-09-438-185-626	Sequence 626, App
36	46	38.7	394	21	US-09-739-449-8822	Sequence 8822, Ap
37	46	38.7	394	22	US-09-803-110-8822	Sequence 8822, Ap
38	46	38.7	476	1	PCT-US01-122636-9	Sequence 9, Appl1
39	46	38.7	476	24	US-60-219-231-9	Sequence 9, Appl1
40	46	38.7	596	16	US-09-252-991A-21255	Sequence 21255, A
41	46	38.7	2590	24	US-60-215-161-7203	Sequence 7203, Ap
42	46	38.7	2603	24	US-60-215-161-5891	Sequence 5891, Ap
43	45	37.8	185	24	US-60-215-161-5719	Sequence 5719, Ap
44	45	37.8	194	1	PCT-US01-14827-9825	Sequence 9825, Ap
45	45	37.8	236	15	US-09-154-750-80	Sequence 80, Appl1

## ALIGNMENTS

RESULT 1  
US-09-290-049-1  
Sequence 1, Application US/09290049  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
FILE REFERENCE: PDC98-0162A  
CURRENT APPLICATION NUMBER: US/09/290,049  
CURRENT FILING DATE: 1999-04-12  
EARLIER APPLICATION NUMBER: 60/081,550  
EARLIER FILING DATE: 1998-04-13  
EARLIER APPLICATION NUMBER: 60/115,142  
EARLIER FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: RAW peptide  
US-09-290-049-1

Query Match 100.0%; Score 119; DB 16; Length 21;  
Best local Similarity 100.0%; Pred. No. 5.3e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ANDHLSTLEAWSNDPTPLD 21  
|||||

Db 1 ANDHLSILEAWSNDPTPYLHD 21

RESULT 2

US-09-290-049-10

Sequence 10, Application US/09290049

GENERAL INFORMATION:

APPLICANT: Smith, Daniel J.

APPLICANT: Taubman, Martin A.

TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL

FILE REFERENCE: FC98-01P2A

CURRENT APPLICATION NUMBER: US/09/290,049

CURRENT FILING DATE: 1999-04-12

EARLIER APPLICATION NUMBER: 60/081,550

EARLIER FILING DATE: 1998-04-13

EARLIER APPLICATION NUMBER: 60/115,142

EARLIER FILING DATE: 1999-01-08

NUMBER OF SEQ ID NOS: 19

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 10

LENGTH: 21

TYPE: PRT

ORGANISM: S. mutans

US-09-290-049-10

Query Match 100.0%; Score 119; DB 16; Length 21;

Best Local Similarity 100.0%; Pred. No. 5.3e-11;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSILEAWSNDPTPYLHD 21

Db 1 ANDHLSILEAWSNDPTPYLHD 21

RESULT 3

US-09-557-848-2

Sequence 2, Application US/09557848

GENERAL INFORMATION:

APPLICANT: Nichols, Scott E.

TITLE OF INVENTION: Substitutes for Modified Starch and

FILE REFERENCE: 0356D2

CURRENT APPLICATION NUMBER: US/09/557,848

CURRENT FILING DATE: 2000-04-26

EARLIER APPLICATION NUMBER: 08/478,704

EARLIER FILING DATE: 1995-06-07

EARLIER APPLICATION NUMBER: 09/007,999

EARLIER FILING DATE: 1998-01-16

NUMBER OF SEQ ID NOS: 2

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 2

LENGTH: 1475

TYPE: PRT

ORGANISM: Streptococcus mutans

US-09-557-848-2

Query Match 100.0%; Score 119; DB 19; Length 1475;

Best Local Similarity 100.0%; Pred. No. 5.2e-09;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSILEAWSNDPTPYLHD 21

Db 481 ANDHLSILEAWSNDPTPYLHD 501

RESULT 4

US-09-740-274-2

Sequence 2, Application US/09740274

GENERAL INFORMATION:

APPLICANT: Nichols, Scott E.

TITLE OF INVENTION: Glucan-containing Compositions and Paper

FILE REFERENCE: 0357CRD

CURRENT APPLICATION NUMBER: US/09/740,274

CURRENT FILING DATE: 2000-12-19

PRIOR APPLICATION NUMBER: 09/210,361

PRIOR FILING DATE: 1998-12-11

PRIOR APPLICATION NUMBER: 09/007,999

PRIOR FILING DATE: 1998-01-16

PRIOR APPLICATION NUMBER: 08/478,704

PRIOR FILING DATE: 1995-06-07

PRIOR APPLICATION NUMBER: 09/009,620

PRIOR FILING DATE: 1998-01-20

PRIOR APPLICATION NUMBER: 08/485,243

PRIOR FILING DATE: 1995-06-07

PRIOR APPLICATION NUMBER: 09/008,172

PRIOR FILING DATE: 1998-01-16

PRIOR APPLICATION NUMBER: 08/482,711

PRIOR FILING DATE: 1995-06-07

NUMBER OF SEQ ID NOS: 6

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 2

LENGTH: 1475

TYPE: PRT

ORGANISM: Streptococcus mutans

US-09-740-274-2

Query Match 100.0%; Score 119; DB 21; Length 1475;

Best Local Similarity 100.0%; Pred. No. 5.2e-09;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSILEAWSNDPTPYLHD 21

Db 481 ANDHLSILEAWSNDPTPYLHD 501

RESULT 5

US-09-740-274-4

Sequence 4, Application US/09740274

GENERAL INFORMATION:

APPLICANT: Nichols, Scott E.

TITLE OF INVENTION: Glucan-containing Compositions and Paper

FILE REFERENCE: 0357CRD

CURRENT APPLICATION NUMBER: US/09/740,274

CURRENT FILING DATE: 2000-12-19

PRIOR APPLICATION NUMBER: 09/210,361

PRIOR FILING DATE: 1998-12-11

PRIOR APPLICATION NUMBER: 09/007,999

PRIOR FILING DATE: 1998-01-16

PRIOR APPLICATION NUMBER: 08/478,704

PRIOR FILING DATE: 1995-06-07

PRIOR APPLICATION NUMBER: 09/009,620

PRIOR FILING DATE: 1998-01-20

PRIOR APPLICATION NUMBER: 08/485,243

PRIOR FILING DATE: 1995-06-07

PRIOR APPLICATION NUMBER: 09/008,172

PRIOR FILING DATE: 1998-01-16

PRIOR APPLICATION NUMBER: 08/482,711

PRIOR FILING DATE: 1995-06-07

NUMBER OF SEQ ID NOS: 6

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 4

LENGTH: 1375

TYPE: PRT

ORGANISM: Streptococcus mutans

US-09-740-274-4

Query Match 92.4%; Score 110; DB 21; Length 1375;

Best Local Similarity 95.2%; Pred. No. 1.2e-07;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANDHLSILEAWSNDPTPYLHD 21

Db 507 ANDHLSILEAMSDNDPTYLHD 527

RESULT 6

US-09-290-049-12

; Sequence 12, Application US/09290049

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.

; APPLICANT: Taubman, Martin A.

; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL

; FILE REFERENCE: FDC98-01P2A

; CURRENT APPLICATION NUMBER: US/09/290,049

; EARLIER FILING DATE: 1999-04-12

; EARLIER APPLICATION NUMBER: 60/081,550

; EARLIER FILING DATE: 1998-04-13

; EARLIER APPLICATION NUMBER: 60/115,142

; EARLIER FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 12

; LENGTH: 21

; TYPE: PRT

; ORGANISM: S. downei

US-09-290-049-12

Query Match

Best Local Similarity 90.8%; Score 108; DB 16; Length 21;

Best Local Similarity 85.7%; Pred. No. 2.7e-09;

Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDPTYLHD 21

DB 1 ANNHVSIVEAMSDNDPTYLHD 21

RESULT 7

US-09-290-049-14

; Sequence 14, Application US/09290049

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.

; APPLICANT: Taubman, Martin A.

; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL

; FILE REFERENCE: FDC98-01P2A

; CURRENT APPLICATION NUMBER: US/09/290,049

; EARLIER FILING DATE: 1999-04-12

; EARLIER APPLICATION NUMBER: 60/081,550

; EARLIER FILING DATE: 1998-04-13

; EARLIER APPLICATION NUMBER: 60/115,142

; EARLIER FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 14

; LENGTH: 21

; TYPE: PRT

; ORGANISM: S. sobrinus

US-09-290-049-14

Query Match

Best Local Similarity 90.8%; Score 108; DB 16; Length 21;

Best Local Similarity 85.7%; Pred. No. 2.7e-09;

Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDPTYLHD 21

DB 1 ANNHVSIVEAMSDNDPTYLHD 21

RESULT 8

US-09-290-049-11

; Sequence 11, Application US/09290049

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.

; APPLICANT: Taubman, Martin A.

; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL

; FILE REFERENCE: FDC98-01P2A

; CURRENT APPLICATION NUMBER: US/09/290,049

; EARLIER FILING DATE: 1999-04-12

; EARLIER APPLICATION NUMBER: 60/081,550

; EARLIER FILING DATE: 1998-04-13

; EARLIER APPLICATION NUMBER: 60/115,142

; EARLIER FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 11

; LENGTH: 21

; TYPE: PRT

; ORGANISM: S. mutans

US-09-290-049-11

Query Match

Best Local Similarity 63.0%; Score 75; DB 16; Length 21;

Best Local Similarity 71.4%; Pred. No. 0.00036;

Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDPTYLHD 21

DB 1 AINHLISLEAMSDNDPQYKND 21

RESULT 9

US-09-649-885-2

; Sequence 2, Application US/09649885

; GENERAL INFORMATION:

; APPLICANT: Nichols, Scott E.

; TITLE OF INVENTION: Substitutes for Modified Starch and

; FILE REFERENCE: 0358D2

; CURRENT APPLICATION NUMBER: US/09/649,885

; EARLIER FILING DATE: 2000-08-28

; EARLIER APPLICATION NUMBER: US 09/008,172

; EARLIER FILING DATE: 1998-01-16

; EARLIER APPLICATION NUMBER: 08/482,711

; EARLIER FILING DATE: 1995-06-07

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 1430

; TYPE: PRT

; ORGANISM: Streptococcus mutans

US-09-649-885-2

Query Match

Best Local Similarity 63.0%; Score 75; DB 20; Length 1430;

Best Local Similarity 71.4%; Pred. No. 0.034;

Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDPTYLHD 21

DB 495 AINHLISLEAMSDNDPQYKND 515

RESULT 10

US-09-740-274-6

; Sequence 6, Application US/09740274

; GENERAL INFORMATION:

; APPLICANT: Nichols, Scott E.

; TITLE OF INVENTION: Glucan-containing Compositions and Paper

; FILE REFERENCE: 0357CRD

; CURRENT APPLICATION NUMBER: US/09/740,274

; EARLIER FILING DATE: 2000-12-19

; EARLIER APPLICATION NUMBER: 09/210,361

; EARLIER FILING DATE: 1998-12-11

```

RESULT 14
US-09-595-298A-660
: Sequence 660, Application US/09595298A
: GENERAL INFORMATION:
: APPLICANT: ALEXANDROV, Nikolai
: ATTORNEY: BROVER, Vyacheslav
: TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypept
: TITLE OF INVENTION: Thereby
: FILE REFERENCE: 2750-0953p
: CURRENT APPLICATION NUMBER: US/09/595,298A
: CURRENT FILING DATE: 2000-06-16
: NUMBER OF SEQ ID NOS: 2756
: SOFTWARE: Patentln version 3.0
: SEQ ID NO 660
: LENGTH: 340
: TYPE: PRT
: ORGANISM: Arabidopsis Thaliana

```



FEATURE:  
: NAME/KEY: peptide  
: LOCATION: (1)..(340)  
: OTHER INFORMATION: Ceres Seq. ID no. 1026740  
: NAME/KEY: misc\_feature  
: LOCATION: ()..()  
: OTHER INFORMATION: Xaa is any aa, unknown or other  
US-09-595-298A-660

Query Match 42.0%; Score 50; DB 19; Length 340;  
Best Local Similarity 52.9%; Pred. No. 55;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 DHTSLTEAWSDNDTPYL 19  
|:|:|||||:|:  
Db 234 DYEQIMEAWSDKGTLYV 250

## RESULT 15

US-60-324-109-17015  
Sequence 17015, Application US/60324109  
GENERAL INFORMATION:

: APPLICANT: Cao, Yongwei  
: APPLICANT: Edgerton, Michael D  
: APPLICANT: Hinkle, Gregory J.  
: APPLICANT: Kovalic, David K.  
: APPLICANT: Liu, Jingdong  
: APPLICANT: Stein, Joshua  
: TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT  
: FILE REFERENCE: 38-10(52726)B  
: CURRENT APPLICATION NUMBER: US/60/324,109  
: CURRENT FILING DATE: 2001-09-21  
: NUMBER OF SEQ ID NOS: 33196  
: SEQ ID NO 17015  
: LENGTH: 546  
: TYPE: PRT  
: ORGANISM: Zea mays  
: FEATURE:  
US-60-324-109-17015

Query Match 41.6%; Score 49.5; DB 24; Length 546;  
Best Local Similarity 58.8%; Pred. No. 1.1e+02;  
Matches 10; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 2 NDHLSILEAWSDNDTPY 18  
|:|:|||||:|:  
Db 144 NDELALETW-DNGKPY 159

Search completed: March 27, 2002, 14:20:24  
Job time: 1571 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:22:46 ; Search time 137.48 seconds  
(without alignments)  
10.540 Million cell updates/sec

Title: US-09-290-049a-10  
Perfect score: 119  
Sequence: 1 ANDHLSILEAWSNDPTLYLHD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 332938 seqs, 6899538 residues

Total number of hits satisfying chosen parameters: 332938

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep:\*  
8: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	21	US-09-562-328-20	Sequence 20, Appl
2	119	100.0	21	US-09-290-049a-1	Sequence 1, Appl
3	119	100.0	21	US-09-290-049a-10	Sequence 10, Appl
4	110	92.4	21	US-09-562-328-21	Sequence 21, Appl
5	108	90.8	21	US-09-562-328-23	Sequence 23, Appl
6	108	90.8	21	US-09-562-328-25	Sequence 25, Appl
7	108	90.8	21	US-09-290-049a-12	Sequence 12, Appl
8	108	90.8	21	US-09-290-049a-14	Sequence 14, Appl
9	75	63.0	21	US-09-562-328-22	Sequence 22, Appl
10	75	63.0	21	US-09-290-049a-11	Sequence 11, Appl
11	75	63.0	21	US-09-604-957-4	Sequence 4, Appl
12	73	61.3	21	US-09-562-328-24	Sequence 24, Appl
13	73	61.3	21	US-09-290-049a-13	Sequence 13, Appl
14	71	59.7	21	US-09-604-957-5	Sequence 5, Appl
15	60	50.4	21	US-09-604-957-6	Sequence 6, Appl
16	49.5	41.6	423	US-09-708-427-55326	Sequence 55326, A
17	49.5	41.6	511	US-09-708-427-55325	Sequence 55325, A
18	49.5	41.6	549	US-09-708-427-55324	Sequence 55324, A
19	49	41.2	302	PCT-US02-03987-14004	Sequence 14004, A
20	49	41.2	302	US-09-815-242-14004	Sequence 14004, A
21	49	41.2	302	US-10-072-851-14004	Sequence 14004, A
22	49	41.2	302	US-09-604-957-7	Sequence 7, Appl
23	49	41.2	1278	US-09-604-957-3	Sequence 3, Appl
24	47.5	39.9	434	US-09-708-427-7994	Sequence 7994, Ap
25	47.5	39.9	519	US-09-708-427-7993	Sequence 7993, Ap

26	47	39.5	15	7	US-10-067-649-90	Sequence 90, Appl
27	47	39.5	366	6	US-09-708-427-27415	Sequence 27415, A
28	47	39.5	367	6	US-09-708-427-27416	Sequence 27416, A
29	47	39.5	378	6	US-09-708-427-27414	Sequence 27414, A
30	46.5	39.1	1604	6	US-09-888-615-73	Sequence 73, Appl
31	46	38.7	199	6	US-09-605-703B-2226	Sequence 2226, Ap
32	46	38.7	2590	6	US-09-897-516-7203	Sequence 7203, Ap
33	46	38.7	2603	6	US-09-897-516-5891	Sequence 5891, Ap
34	45	37.8	185	6	US-09-897-516-5719	Sequence 5719, Ap
35	45	37.8	236	6	US-09-154-750A-80	Sequence 80, Appl
36	45	37.8	638	6	US-09-708-427-25322	Sequence 25322, A
37	45	37.8	765	6	US-09-708-427-25321	Sequence 25321, A
38	45	37.8	947	6	US-09-708-427-25320	Sequence 25320, A
39	44.5	37.4	499	6	US-09-614-150-42525	Sequence 42525, A
40	44	37.0	322	1	PCT-US02-03987-11634	Sequence 11634, A
41	44	37.0	322	1	US-09-815-242-11634	Sequence 11634, A
42	44	37.0	322	7	US-10-072-851-11634	Sequence 11634, A
43	44	37.0	331	1	PCT-US02-03987-10188	Sequence 10188, A
44	44	37.0	331	6	US-09-815-242-10188	Sequence 10188, A
45	44	37.0	331	7	US-10-072-851-10188	Sequence 10188, A

#### ALIGNMENTS

RESULT 1  
US-09-562-328-20  
; Sequence 20, Application US/09562328  
; GENERAL INFORMATION:  
; APPLICANT: LEES, ANDREW  
; APPLICANT: TAUBMAN, MARTIN A.  
; APPLICANT: SMITH, DANIEL J.  
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARRIES  
; FILE REFERENCE: 04995.0046-01  
; CURRENT FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 09/288,965  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 20  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Streptococcus sp.  
US-09-562-328-20

Query Match 100.0%; Score 119; DB 6; Length 21;  
Best Local Sim. Early 100.0%; Pred. No. 2.9e-12;  
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ANDHLSILEAWSNDPTLYLHD 21  
Db 1 ANDHLSILEAWSNDPTLYLHD 21  
RESULT 2  
US-09-290-049a-1  
; Sequence 1, Application US/09290049A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: EAW peptide  
US-09-290-049A-1

Query Match 100.0%; Score 119; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2,9e-12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHSTLEAMSNDTPYLHD 21  
DB 1 ANDHSTLEAMSNDTPYLHD 21

US-09-290-049A-10

Sequence 10, Application US/09290049A  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
FILE REFERENCE: 1564,1008-002  
CURRENT APPLICATION NUMBER: US/09/290,049A  
CURRENT FILING DATE: 1999-04-12  
PRIOR APPLICATION NUMBER: 60/081,550  
PRIOR FILING DATE: 1998-04-13  
PRIOR APPLICATION NUMBER: 60/115,142  
PRIOR FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10  
LENGTH: 21  
TYPE: PRT  
ORGANISM: S. mutans  
US-09-290-049A-10

Query Match 100.0%; Score 119; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2,9e-12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ANDHSTLEAMSNDTPYLHD 21  
DB 1 ANDHSTLEAMSNDTPYLHD 21

US-09-562-328-21

Sequence 21, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 04995,0046-01  
CURRENT APPLICATION NUMBER: US/09/562,328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 21  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-21

Query Match 92.4%; Score 110; DB 6; Length 21;  
Best Local Similarity 95.2%; Pred. No. 7,4e-11;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ANDHSTLEAMSNDTPYLHD 21  
DB 1 ANDHSTLEAMSNDTPYLHD 21

US-09-562-328-23

Sequence 23, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 04995,0046-01  
CURRENT APPLICATION NUMBER: US/09/562,328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 23  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-23

Query Match 90.8%; Score 108; DB 6; Length 21;  
Best Local Similarity 85.7%; Pred. No. 1,5e-10;  
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHSTLEAMSNDTPYLHD 21  
DB 1 ANNHVSIVEAMSNDTPYLHD 21

US-09-562-328-25

Sequence 25, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 04995,0046-01  
CURRENT APPLICATION NUMBER: US/09/562,328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 25  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-25

Query Match 90.8%; Score 108; DB 6; Length 21;  
Best Local Similarity 85.7%; Pred. No. 1,5e-10;  
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHSTLEAMSNDTPYLHD 21  
DB 1 ANNHVSIVEAMSNDTPYLHD 21

US-09-290-049A-12

```
; Sequence 12, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290.049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 21
; TYPE: PRF
; ORGANISM: S. downei
US-09-290-049A-12

Query Match          90.8%; Score 108; DB 6; Length 21;
Best Local Similarity 85.7%; Pred. No. 1.5e-10;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHSLLEAMSDNDTPYLHD 21
   ||:|||||:|||||:|||||
Db 1 AINHSLLEAMSDNDTPYLHD 21

RESULT 8
US-09-290-049A-14
; Sequence 14, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290.049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 21
; TYPE: PRF
; ORGANISM: S. sobrinus
US-09-290-049A-14

Query Match          90.8%; Score 108; DB 6; Length 21;
Best Local Similarity 85.7%; Pred. No. 1.5e-10;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHSLLEAMSDNDTPYLHD 21
   ||:|||||:|||||:|||||
Db 1 AINHSLLEAMSDNDTPYLHD 21

RESULT 9
US-09-562-328-22
; Sequence 22, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEES, ANDREW
; APPLICANT: TAUBMAN, MARTIN A.
; APPLICANT: SMITH, DANIEL J.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
; FILE REFERENCE: 04995.0046-01
```

```
; CURRENT APPLICATION NUMBER: US/09/562.328
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 09/288,965
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 21
; TYPE: PRF
; ORGANISM: Streptococcus sp.
US-09-562-328-22

Query Match          63.0%; Score 75; DB 6; Length 21;
Best Local Similarity 71.4%; Pred. No. 2.3e-05;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANDHSLLEAMSDNDTPYLHD 21
   ||:|||||:|||||:|||||
Db 1 AINHSLLEAMSDNDTPYLHD 21

RESULT 10
US-09-290-049A-11
; Sequence 11, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290.049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 21
; TYPE: PRF
; ORGANISM: S. mutans
US-09-290-049A-11

Query Match          63.0%; Score 75; DB 6; Length 21;
Best Local Similarity 71.4%; Pred. No. 2.3e-05;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANDHSLLEAMSDNDTPYLHD 21
   ||:|||||:|||||:|||||
Db 1 AINHSLLEAMSDNDTPYLHD 21

RESULT 11
US-09-604-957-6
; Sequence 4, Application US/09604957
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUTI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604.957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 545
```

TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-604-957-4

Query Match 63.0%; Score 75; DB 6; Length 545;  
Best Local Similarity 71.4%; Pred. No. 0.00094;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANDHLSILEAWSNDNPTPLHD 21  
| : ||||| ||||| || :  
Db 75 AINHLSTLEAWSNDNPQYVKD 95

RESULT 12  
US-09-562-328-24  
Sequence 24, Application US/09562328

GENERAL INFORMATION:

APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
APPLICANT: SMITH, DANIEL J.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 04995.0046-01  
CURRENT APPLICATION NUMBER: US/09/562.328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288.965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 24  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-24

Query Match 61.3%; Score 73; DB 6; Length 21;  
Best Local Similarity 73.7%; Pred. No. 4.7e-05;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANDHLSILEAWSNDNPTPL 19  
| : ||||| ||||| || :  
Db 1 AIDHLSILEAWSGNDNDYV 19

RESULT 13  
US-09-290-049a-13

Sequence 13, Application US/09290049a

GENERAL INFORMATION:

APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
FILE REFERENCE: 1564.1008-002  
CURRENT APPLICATION NUMBER: US/09/290.049a  
CURRENT FILING DATE: 1999-04-12  
PRIOR APPLICATION NUMBER: 60/081.550  
PRIOR FILING DATE: 1998-04-13  
PRIOR APPLICATION NUMBER: 60/115.142  
PRIOR FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13  
LENGTH: 21  
TYPE: PRT  
ORGANISM: S. downei  
US-09-290-049a-13

Query Match 61.3%; Score 73; DB 6; Length 21;  
Best Local Similarity 73.7%; Pred. No. 4.7e-05;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANDHLSILEAWSNDNPTPL 19  
| : ||||| ||||| || :  
Db 1 AIDHLSILEAWSGNDNDYV 19

RESULT 14  
US-09-604-957-5  
Sequence 5, Application US/09604957

GENERAL INFORMATION:

APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
APPLICANT: DIKSHUIZEN, LUBBERT  
APPLICANT: RAHAOUI, HAKIM  
APPLICANT: LEER, ROBERT-JAN  
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
FILE REFERENCE: BO 43388  
CURRENT APPLICATION NUMBER: US/09/604.957  
CURRENT FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 523  
TYPE: PRT  
ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-5

Query Match 59.7%; Score 71; DB 6; Length 523;  
Best Local Similarity 66.7%; Pred. No. 0.0038;  
Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANDHLSILEAWSNDNPTPLHD 21  
| : ||||| ||||| || :  
Db 75 ANOHLSTLEAWSNDNPLXYTD 95

RESULT 15  
US-09-604-957-6

Sequence 6, Application US/09604957

GENERAL INFORMATION:

APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
APPLICANT: DIKSHUIZEN, LUBBERT  
APPLICANT: RAHAOUI, HAKIM  
APPLICANT: LEER, ROBERT-JAN  
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
FILE REFERENCE: BO 43388  
CURRENT APPLICATION NUMBER: US/09/604.957  
CURRENT FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 6  
LENGTH: 584  
TYPE: PRT  
ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-6

Query Match 50.4%; Score 60; DB 6; Length 584;  
Best Local Similarity 55.0%; Pred. No. 0.23;  
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANDHLSILEAWSNDNPTPLHD 20  
| : ||||| ||||| || :  
Db 75 ANKHLSTLEAWSNDNPQYVYN 94

Search completed: March 27, 2002, 14:22:47  
Job time: 1694 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:01:17 ; Search time 102.51 Seconds

(without alignments)  
15.605 Million cell updates/sec

Title: US-09-290-049a-10

Perfect score: 119

Sequence: 1 ANDHLSILEAWSNDNDPYLHD 21

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	1475	2 B33135	gtfB protein precu
2	110	92.4	1375	2 UT0345	dextranucrase (EC
3	108	90.8	1592	2 A38175	glucosyltransferas
4	78	65.0	1365	2 A41483	glucosyltransferas
5	75	63.0	1431	2 A45866	dextranucrase (EC
6	74	62.2	1508	2 T31098	probable dextranu
7	66	55.5	1449	2 T30857	glucosyltransferas
8	66	55.5	1449	2 T30552	glucosyltransferas
9	63	52.9	1577	2 T30858	glucosyltransferas
10	58	48.7	1518	2 A44811	glucosyltransferas
11	57	47.9	331	2 B48445	glycerolaldehyde-3-p
12	57	47.9	1599	2 S22737	glucosyltransferas
13	52	43.7	4848	2 T30289	glucosyltransferas
14	50	42.0	347	2 T48610	pistinnamycin I sy
15	49.5	41.6	549	2 T03983	hypothetical prote
16	48	40.3	336	1 H0538	r12 nuclear restor
17	48	40.3	490	2 H0538	BGLF2 protein - hu
18	47.5	39.9	331	1 DEUNGC	probable ppdk prot
19	47.5	39.9	519	2 C86372	glycerolaldehyde-3-p
20	47	39.5	175	2 C86205	hypothetical prote
21	47	39.5	378	2 T04254	hypothetical prote
22	47	39.5	524	2 D82220	conserved hypotet
23	47	39.5	525	2 T40088	RhoGEF domain cont
24	46	38.7	275	2 S55978	hypothetical prote
25	46	38.7	335	2 B86568	glycerolaldehyde-3-p
26	46	38.7	335	2 B72053	glycerolaldehyde-3-p
27	46	38.7	335	2 S58229	salicylate biosynt
28	45.5	38.2	96	2 E81786	conserved hypotet
29	45	37.8	236	2 S62732	guanidinacetate N

30	45	37.8	947	2 E86362	hypothetical prote
31	45	37.8	1159	2 T43461	probable phosphodi
32	44.5	37.4	210	2 H83332	conserved hypotet
33	44	37.0	247	2 PQ0178	glycerolaldehyde-3-p
34	44	37.0	331	2 DEECG3	glycerolaldehyde-3-p
35	44	37.0	331	2 D85788	glycerolaldehyde-3-p
36	44	37.0	333	2 DECHG3	glycerolaldehyde-3-p
37	44	37.0	333	2 JC5370	glycerolaldehyde-3-p
38	44	37.0	337	1 DEJUC	glycerolaldehyde-3-p
39	44	37.0	337	1 DEZMCC	glycerolaldehyde-3-p
40	44	37.0	337	1 DEBHG	glycerolaldehyde-3-p
41	44	37.0	436	2 T24953	hypothetical prote
42	44	37.0	468	2 S61964	probable membrane
43	44	37.0	1208	2 B82091	exodeoxyribonuclea
44	44	37.0	1313	2 T29027	hypothetical prote
45	43.5	36.6	97	2 S12958	tachyinnin delta p

#### ALIGNMENTS

```

RESULT 1
B33135
gtfB protein precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 15-Oct-1999
C:Accession: B33135; A33128
R:Shiroza, T.; Ueda, S.; Kuramitsu, H. K.
J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A:Reference number: A33135; MUID:87308013
A:Accession: B33135
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1475 <SH1>
A:Cross-references: GB:M17361; NID:G153639; PIDN:AAA8588.1; PID:G153640
R:Shiroza, T.; Ueda, S.; Kuramitsu, H. K.
submitted to the Protein Sequence Database, September 1990
A:Reference number: A33128
A:Accession: A33128
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-171, 173-641, 'N', 643-1475 <SH2>
A:Experimental source: strain GS-5
C:Superfamily: cpl repeat homology
F:1096-1115/Domain: cpl repeat homology <CP1>
F:1224-1243/Domain: cpl repeat homology <CP2>
F:1289-1308/Domain: cpl repeat homology <CP3>
F:1354-1373/Domain: cpl repeat homology <CP4>
F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match 100.0%; Score 119; DB 2; Length 1475;
Best Local Similarity 100.0%; Pred. No. 3.9e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDNDPYLHD 21
DB 481 ANDHLSILEAWSNDNDPYLHD 501

RESULT 2
UT0345
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)
N:Alternate names: sucrose 6-glucosyltransferase
C:Species: Streptococcus mutans
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 24-Sep-1999
C:Accession: UT0345; C33135
R:Ueda, S.; Shiroza, T.; Kuramitsu, H. K.
Gene 69, 101-109, 1988
A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.
A:Reference number: UT0345; MUID:89137980
A:Accession: UT0345

```

A:Molecule type: DNA  
A:Residues: 1-1375 <UED>  
A:Experimental source: GS-5  
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
J. Bacteriol. 169, 4263-4270, 1987  
A:Title: Sequence analysis of the *glfB* gene from *Streptococcus mutans*.  
A:Reference number: A3135; MUID:87308013  
A:Accession: C3135  
A:Molecule type: DNA  
A:Residues: 1-349 <SHI>  
A:Cross-references: GB:M17361  
C:Geneclis:  
A:Gene: *glfC*  
C:Function:  
A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans  
C:Superfamily: cpl repeat homology  
Keywords: duplication; glycosyltransferase; hexosyltransferase  
-34/Domain: signal sequence #status predicted <Sig>  
-35-1375/Product: glycosyltransferase #status predicted <MAT>  
F:1126-1145/Domain: cpl repeat homology <CP1>  
F:1253-1272/Domain: cpl repeat homology <CP2>  
F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 92.4%; Score 110; DB 2; Length 1375;  
Best Local Similarity 95.2%; Pred. No. 9, 1e-09;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ANDHSLILEAWSNDNPTPLHD 21  
DB 507 ANDHSLILEAWSNDNPTPLHD 527

RESULT 3  
A38175  
glucosyltransferase precursor - *Streptococcus sobrinus*  
C:Species: *Streptococcus sobrinus*  
C:Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 15-Oct-1999  
C:Accession: A38175  
R:Abo, H.; Mtsunura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.  
J. Bacteriol. 173, 989-996, 1991  
A:Title: Peptide sequences for sucrose splitting and glucan binding within *Streptococcus*  
A:Reference number: A38175; MUID:91123227  
A:Accession: A38175  
A:Status: preliminary  
Molecule type: DNA  
Residues: 1-1592 <ABO>  
A:Cross-references: GB:D90213; NID:9217032; PIDN:BAAL4241.1; PID:d1014946; PID:9217033  
C:Superfamily: cpl repeat homology  
F:1093-1112/Domain: cpl repeat homology <CP1>  
F:1222-1241/Domain: cpl repeat homology <CP2>  
F:1287-1306/Domain: cpl repeat homology <CP3>  
F:1330-1351/Domain: cpl repeat homology <CP4>  
F:1352-1371/Domain: cpl repeat homology <CP5>  
F:1402-1420/Domain: cpl repeat homology <CP6>  
F:1465-1484/Domain: cpl repeat homology <CP7>  
F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 90.8%; Score 108; DB 2; Length 1592;  
Best Local Similarity 85.7%; Pred. No. 2, 2e-08;  
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHSLILEAWSNDNPTPLHD 21  
DB 477 ANNHVSIVEAWSNDNPTPLHD 497

RESULT 4  
A41483  
glucosyltransferase (EC 2.4.1.-) *glfS* precursor - *Streptococcus sobrinus*  
C:Species: *Streptococcus sobrinus*

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999  
C:Accession: A41483  
R:Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.  
Infect. Immun. 58, 2452-2458, 1990  
A:Title: Analysis of the *Streptococcus downei glfS* gene, which specifies a glucosyltr  
A:Reference number: A41483; MUID:90316665  
A:Accession: A41483  
A:Molecule type: DNA  
A:Residues: 1-1365 <GIL>  
A:Cross-references: GB:M30943; NID:9153652; PIDN:AAA26898.1; PID:9153653  
C:Geneclis:  
A:Gene: *glfS*  
C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 65.5%; Score 78; DB 2; Length 1365;  
Best Local Similarity 71.4%; Pred. No. 0.00087;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 ANDHSLILEAWSNDNPTPLHD 21  
DB 467 AIDHSLILEAWSNDNPTPLHD 487

RESULT 5  
A45866  
dextranucrase (EC 2.4.1.5) precursor - *Streptococcus mutans*  
C:Species: *Streptococcus mutans*  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: A45866  
R:Honda, O.; Kato, C.; Kuramitsu, H.K.  
J. Gen. Microbiol. 136, 2099-2105, 1990  
A:Title: Nucleotide sequence of the *Streptococcus mutans glfD* gene encoding the gluco  
A:Reference number: A45866; MUID:91100958  
A:Accession: A45866  
A:Status: preliminary  
Molecule type: DNA  
Residues: 1-1431 <HON>  
A:Cross-references: GB:M29296  
C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:181-201/Domain: cpl repeat homology <CP1>  
F:1127-1146/Domain: cpl repeat homology <CP2>  
F:1192-1211/Domain: cpl repeat homology <CP3>  
F:1257-1276/Domain: cpl repeat homology <CP4>  
F:1327-1397/Domain: cpl repeat homology <CP5>  
F:1321-1340/Domain: cpl repeat homology <CP6>  
F:1341-1361/Domain: cpl repeat homology <CP7>  
F:1385-1404/Domain: cpl repeat homology <CP8>

Query Match 63.0%; Score 75; DB 2; Length 1431;  
Best Local Similarity 71.4%; Pred. No. 0.0027;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 ANDHSLILEAWSNDNPTPLHD 21  
DB 495 AINHLILEAWSNDNPTPLHD 515

RESULT 6  
T31098  
probable dextranucrase (EC 2.4.1.5), extracellular - *Leuconostoc mesenteroides*  
C:Species: *Leuconostoc mesenteroides*  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
C:Accession: T31098  
R:Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willems, R.M.  
FEBS Microbiol. Lett. 159, 307-315, 1998  
A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase  
A:Reference number: Z20981; MUID:98164374  
A:Accession: T31098  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
A:Residues: 1-1508 <MON>  
A:Cross-references: EMBL:AF030129; NID:g2766611; PID:g2766612; PIDN:AAB95453.1  
A:Experimental source: strain NRRL B-1299  
C:Genetics:  
A:Gene: dsrB  
C:Function:  
A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds  
C:Keywords: glycosyltransferase, hexosyltransferase

Query Match 62.2%; Score 74; DB 2; Length 1508;  
Best Local Similarity 66.7%; Pred. No. 0.0041;  
Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANDHSLTLEAMSDNDTPYLHD 21  
||| ||||| ||| ||| |||  
Db 563 ANQHLSTLEDMWSHNDPEYKVD 583

RESULT 7  
857  
glycosyltransferase - Streptococcus salivarius

C:Species: Streptococcus salivarius  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30857  
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
Infect. Immun. 63, 609-621, 1995  
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pr  
A:Reference number: 220909; MUID:95122197  
A:Accession: T30857  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1449 <SIM>  
A:Cross-references: EMBL:L35495; NID:g662378; PID:g662379; PIDN:AAC41412.1  
C:Genetics:  
A:Gene: gtfI

Query Match 55.5%; Score 66; DB 2; Length 1449;  
Best Local Similarity 72.2%; Pred. No. 0.069;  
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 HLSLEAMSDNDTPYLHD 21  
||||| ||| ||| |||  
Db 539 HLSLEAMSHNDAYNED 556

RESULT 8  
10552  
glycosyltransferase N - Streptococcus salivarius (fragment)

C:Species: Streptococcus salivarius  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30552  
R:Jaffe, R.I.  
submitted to the EMBL Data Library, February 1998  
A:Description: Streptococcus salivarius VI477 gtfN.  
A:Reference number: 220854  
A:Accession: T30552  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1449 <JAF>  
A:Cross-references: EMBL:AF049609; NID:g2935545; PID:g2935546; PIDN:AAC05156.1  
C:Genetics:  
A:Gene: gtfN

Query Match 55.5%; Score 66; DB 2; Length 1449;  
Best Local Similarity 72.2%; Pred. No. 0.069;  
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 4 HLSLEAMSDNDTPYLHD 21  
||||| ||| ||| |||

Db 539 HLSLEAMSHNDAYNED 556

RESULT 9  
T30858  
glycosyltransferase - Streptococcus salivarius

C:Species: Streptococcus salivarius  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30858  
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
Infect. Immun. 63, 609-621, 1995  
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for

A:Reference number: 220909; MUID:95122197  
A:Accession: T30858  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
A:Residues: 1-1577 <SIM>  
A:Cross-references: EMBL:L35928; NID:g662380; PID:g662381; PIDN:AAC41413.1  
C:Genetics:  
A:Gene: gtfm

Query Match 52.9%; Score 63; DB 2; Length 1577;  
Best Local Similarity 72.2%; Pred. No. 0.23;  
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 HLSLEAMSDNDTPYLHD 21  
||||| ||| ||| |||  
Db 591 HLSLEAMSYNDHQYKND 608

RESULT 10

A44811  
glycosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 15-Oct-1999  
C:Accession: A44811; S22726; S28809  
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
J. Gen. Microbiol. 137, 2577-2593, 1991  
A:Title: Molecular characterization of a cluster of at least two glycosyltransferase  
A:Reference number: A44811; MUID:92148377  
A:Accession: A44811  
A:Molecule type: DNA  
A:Residues: 1-1518 <GTF>  
A:Cross-references: EMBL:Z11873; NID:g47526; PIDN:CAAT7900.1; PID:g47527  
A:Note: sequence extracted from NCBI Backbone (NCBIN:81050, NCBIIP:81052)  
C:Genetics:  
A:Gene: gtfJ  
C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase, hexosyltransferase  
F:1307-1326/Domain: cpl repeat homology <CP4>

Query Match 48.7%; Score 58; DB 2; Length 1518;  
Best Local Similarity 66.7%; Pred. No. 1.3;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 HLSLEAMSDNDTPYLHD 18  
||| ||||| ||| ||| |||  
Db 504 HSLVLEAMSLNDNH 518

RESULT 11

B48445  
glyceralddehyde-3-phosphate dehydrogenase (EC 1.2.1.12) - Leishmania mexicana  
C:Species: Leishmania mexicana  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 11-Jun-1999  
C:Accession: B48445; S25142  
R:Hannaert, V.; Blaauw, M.; Kohl, L.; Allert, S.; Opperdoes, F.R.; Michels, P.A.M.  
Mol. Biochem. Parasitol. 55, 115-126, 1992  
A:Title: Molecular analysis of the cytosolic and glycosomal glyceralddehyde-3-phosphat  
A:Reference number: A48445; MUID:93063042  
A:Accession: B48445

A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-331 <HAN>  
 A:Cross-references: EMBL:X65220; NID:g9552; PIDN:CAA6323.1; PID:g9553  
 C:Superfamily: glyceroldehyde-3-phosphate dehydrogenase  
 C:Keywords: oxidoreductase

Query Match 47.9%; Score 57; DB 2; Length 331;  
 Best Local Similarity 52.6%; Pred. No. 0.31;  
 Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 NDHLSLEAWSNDTPYLH 20  
 DB 301 NDHFKLVSMYDNETGYSH 319

SULT 12  
 2737

Glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius

C:Species: Streptococcus salivarius  
 C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000  
 C:Accession: S22737; S28810; B4811; S22727

R:Jacques, N.

submitted to the EMBL Data Library, March 1992

A:Reference number: S22726

C:Accession: S22737

A:Molecule type: DNA

A:Residues: 1-1599 <JAC>

A:Cross-references: EMBL:Z11872; NID:g47530; PIDN:CAA77998.1; PID:g47531

A:Experimental source: ATCC 25975

R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.

J. Gen. Microbiol. 137, 2577-2593, 1991

A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes

A:Reference number: A44811; MUID:92148377

C:Accession: S28810

A:Molecule type: DNA

A:Residues: 1-51 <GIF>

A:Cross-references: EMBL:Z11873

C:Genetics:

A:Gene: gtfK

C:Superfamily: cpl repeat homology

C:Keywords: glycosyltransferase; hexosyltransferase

F:1456-1475/Domain: cpl repeat homology <CPR>

Query Match 47.9%; Score 57; DB 2; Length 1599;  
 Best Local Similarity 55.6%; Pred. No. 2;  
 Matches 10; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 4 HSLSLAWSNDTPYLHD 21  
 DB 494 NISLEAWSNDTPYVNE 511

RESULT 13

T30289

pristinamycin I synthase 3 - Streptomyces pristinaespiralis

C:Species: Streptomyces pristinaespiralis

C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 01-Dec-2000

C:Accession: T30289

R:de Crecy-Lagard, V.A.; Saurin, W.; Thibaut, D.; Gil, P.; Naudin, L.; Crouzet, J.; Bian

submitted to the EMBL Data Library, February 1997

A:Description: Streptogramin B biosynthesis in Streptomyces pristinaespiralis and Strept

A:Reference number: Z20808

C:Accession: T30289

A:Molecule type: translated from GB/EMBL/DBJ

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-4848 <DEC>

A:Cross-references: EMBL:Y11548; NID:e1025755; PID:e307539; PIDN:CAA72312.1

C:Genetics:

A:Gene: sbdE

C:Superfamily: acyl carrier protein homology; acetate-CoA ligase homology

C:Keywords: carrier protein; phosphopantetheine; phosphoprotein

F:511-951/Domain: acetate--CoA ligase homology <ACL1>

F:966-1034/Domain: acyl carrier protein homology <ACP1>

F:1563-2024/Domain: acetate--CoA ligase homology #status: atypical <ACL2>

F:2399-2467/Domain: acyl carrier protein homology <ACP2>

F:2995-3427/Domain: acetate--CoA ligase homology <ACP3>

F:3441-3509/Domain: acyl carrier protein homology <ACP3>

F:4043-4492/Domain: acetate--CoA ligase homology <ACP4>

F:4507-4575/Domain: acyl carrier protein homology <ACP4>

F:998,2431,3473/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 43.7%; Score 52; DB 2; Length 4848;  
 Best Local Similarity 47.1%; Pred. No. 44;  
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANDHLSLEAWSNDTP 17  
 DB 3972 ADDHARVLGWNDDHP 3988

RESULT 14

T48610

hypothetical protein F18022.160 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T48610

R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.

submitted to the Protein Sequence Database, April 2000

A:Reference number: Z24493

C:Accession: T48610

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-347 <BEV>

A:Cross-references: EMBL:DB

A:Experimental source: cultivar Columbia; BAC clone F18022

C:Genetics:

A:Map position: 5

A:Introns: 232/3; 296/3

A:Note: F18022.160

Query Match 42.0%; Score 50; DB 2; Length 347;  
 Best Local Similarity 52.9%; Pred. No. 4;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 DHLSLEAWSNDTPYL 19  
 DB 200 DYEQIMEAWSDKCTLYV 216

RESULT 15

T03983

rf2 nuclear restorer protein - maize

N:Alternate names: aldehyde dehydrogenase homolog

C:Species: Zea mays (maize)

C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000

C:Accession: T03983

R:Cui, X.; Wise, R.P.; Schnable, P.S.

Science 272, 1334-1336, 1996

A:Title: The rf2 nuclear restorer gene of male-sterile T-cytoplasm maize.

A:Reference number: Z15177; MUID:96243131

C:Accession: T03983

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-349 <CU1>

A:Cross-references: EMBL:U43082; NID:g1421729; PIDN:AAC49371.1; PID:g1421730

A:Experimental source: strain B73

C:Genetics:

A:Map position: 9

A:Note: rf2

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:107-370/Domain: aldehyde dehydrogenase homology <ALD>

Query Match 41.6%; Score 49.5; DB 2; Length 549;  
Best Local Similarity 58.8%; Pred. No. 8.2;  
Matches 10; Conservative 1; Mismatches 5; Indels 1; Gaps 1;  
QY 2 NDHLSILEAWSDNDTPY 18  
|||:|||||  
Db 147 NDELALETW-DNGKPY 162

Search completed: March 27, 2002, 14:01:18  
Job time: 480 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:27:02 ; Search time 53.4 Seconds  
(without alignments)  
14.419 Million cell updates/sec

Title: US-09-290-049a-10  
Perfect score: 119  
Sequence: 1 ANDHLSILEAMSDNDPVLHD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues  
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SWISSProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	1476	1	GTFB_STRMU
2	110	92.4	1375	1	GTFB_STRMU
3	108	90.8	1592	1	GTF2_STRDO
4	108	90.8	1597	1	GTF1_STRDO
5	78	65.5	1365	1	GTFB_STRDO
6	75	63.0	1462	1	GTFD_STRDO
7	57	47.9	330	1	G3PC_LEIME
8	49	41.2	336	1	UL16_EBV
9	47.5	39.9	330	1	G3PC_TRYB
10	46	38.7	275	1	YG33_YEAST
11	46	38.7	335	1	G3P_CHIPN
12	46	38.7	476	1	PCHA_PSEAE
13	45	37.8	236	1	GAMT_HUMAN
14	45	37.8	385	1	GUNE_FUSOX
15	44.5	37.4	470	1	AMP2_STRCO
16	44	37.0	234	1	G3P_MEIGA
17	44	37.0	330	1	G3PL_ECOLI
18	44	37.0	332	1	G3P_CHICK
19	44	37.0	332	1	G3P_COLLI
20	44	37.0	337	1	G3PC_MAIZE
21	44	37.0	337	1	G3PD_MAIZE
22	44	37.0	337	1	G3PX_HORVU
23	44	37.0	337	1	G3P_CRYPA
24	44	37.0	363	1	G3PL_JACOR
25	44	37.0	816	1	SUS2_MAIZE
26	43.5	36.6	115	1	TKN1_RABIT
27	43.5	36.6	130	1	TKN1_MESAU
28	43.5	36.6	130	1	TKN1_MOUSE
29	43.5	36.6	130	1	TKN1_RAT
30	43.5	36.6	491	1	G6PD_ERWCH
31	43	36.1	235	1	GAMT_RAT
32	43	36.1	236	1	GAMT_MOUSE
33	43	36.1	320	1	G3P_BOVIN

34	43	6.1	322	1	G3P_SHEEP
35	43	6.1	332	1	G3P_PIG
36	43	36.1	332	1	G3P_RABIT
37	43	36.1	334	1	G3P2_HUMAN
38	43	36.1	337	1	G3P_MYGE
39	43	36.1	424	1	IVD_RAT
40	43	36.1	674	1	COAA_BACTJ
41	43	36.1	2485	1	PTND_HUMAN
42	42.5	35.7	441	1	HGD_STRCO
43	42.5	35.7	491	1	G6PD_ECOLI
44	42.5	35.7	645	1	IME2_YEAST
45	42.5	35.7	764	1	Y1S7_YEAST

## ALIGNMENTS

RESULT 1  
ID GTFB\_STRMU STANDARD: PRT: 1476 AA.  
AC P08987; 069381; 069384; 069387; 069390; 069396;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)  
DE (SUCCROSE 6-GLUCOSYLTRANSFERASE).  
GN GTFB.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-GS-5;  
RX MEDLINE=87308013; PubMed=3040685;  
RA Shiroza T., Ueda S., Kuramitsu H.K.;  
RT "Sequence analysis of the gtfb gene from Streptococcus mutans.";  
RL J. Bacteriol. 169:4263-4270(1987).  
[2]  
RN SEQUENCE FROM N.A.  
RC STRAIN-MT4239, MT4245, MT4251, MT4467, AND MT8148;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,  
Kimura S., Hamada S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans.";  
RL FEBS Microbiol. Lett. 161:331-336(1998).  
CC -I- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -I- CATALYTIC ACTIVITY: SUCCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) = D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).  
CC -I- SUBCELLULAR LOCATION: SECRETED.  
CC -I- DISEASE: DENTAL CARIES.  
CC -I- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF GLUCANS.  
CC -I- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-BINDING PROTEIN FROM S. MUTANS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC EMBL: M17361; AAA88588.1;  
CC EMBL: D88651; BAA26101.1;  
CC EMBL: D88654; BAA26105.1;  
DR

```

DR EMBL: D88657; BAA26109.1; -
DR EMBL: D88660; BAA26113.1; -
DR EMBL: D89977; BAA26119.1; -
DR PIR: B3135; B3135.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1.13.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Transferase: Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 34
FT CHAIN 35 1476
FT DOMAIN 35 1051
FT DOMAIN 1097 1476
FT DOMAIN 1097 1130
FT DOMAIN 1161 1470
FT REPEAT 1161 1210
FT REPEAT 1225 1275
FT REPEAT 1290 1340
FT REPEAT 1355 1405
FT REPEAT 1420 1470
FT REPEAT 62 62
FT VARIANT 62 65
FT VARIANT 65 65
FT VARIANT 68 68
FT VARIANT 78 78
FT VARIANT 86 86
FT VARIANT 89 89
FT VARIANT 168 168
FT VARIANT 276 276
FT VARIANT 399 399
FT VARIANT 474 474
FT VARIANT 512 512
FT VARIANT 519 519
FT VARIANT 701 701
FT VARIANT 708 708
FT VARIANT 938 938
FT VARIANT 952 957
FT VARIANT 963 964
FT VARIANT 968 970
FT VARIANT 1086 1086
FT VARIANT 1158 1158
FT VARIANT 1163 1163
FT VARIANT 1168 1168
FT VARIANT 1182 1182
FT VARIANT 1234 1234
FT VARIANT 1263 1263
FT VARIANT 1263 1263
FT VARIANT 1264 1264
FT VARIANT 1272 1272
FT VARIANT 1329 1329
FT VARIANT 1394 1394
FT VARIANT 1402 1402
FT VARIANT 1459 1459
FT CONFLICT 570 570
FT CONFLICT 800 817
FT CONFLICT 1310 1310
FT CONFLICT 1476 AA; 165685 MW; 34798628076549D98 CRC64;
SQ SEQUENCE 1476 AA; 165685 MW; 34798628076549D98 CRC64;

Query Match 100.0%; Score 119; DB 1; Length 1476;
Best Local Similarity 100.0%; Pred. No. 2.1e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ANDHLSLEAMSNDTPYLYD 21
DB 481 ANDHLSLEAMSNDTPYLYD 501

```

```

RESULT 2
ID GTFC_STRMU STANDARD; PRT; 1375 AA.
AC P13470; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)
DE (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).
DE GTF.
GN Streptococcus mutans.
OS Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxId=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GS-5;
RX MEDLINE=89137980; PubMed=2976010;
RA Ueda S., Shiroza T., Kuramitsu H.K.;
RT "Sequence analysis of the glfC gene from Streptococcus mutans GS-5.";
RL Gene 69:101-109(1988).
RN [2]
RP SEQUENCE OF 1-349 FROM N.A.
RC STRAIN-GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the glfB gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) =
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- MISCELLANEOUS: GTF-SI SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-SI SYNTHESIZES BOTH
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: M22054; AAA88592.1; -
DR EMBL: M17361; AAA88589.1; -
DR PIR: J0345; J0345.
DR PIR: C3135; C3135.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1.7.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Transferase: Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 34
FT CHAIN 35 1375
FT DOMAIN 35 1050
FT DOMAIN 1126 1375
FT DOMAIN 1126 1375
FT REPEAT 1126 1159
FT REPEAT 1169 1200
FT REPEAT 1227 1238
FT REPEAT 1253 1303
FT REPEAT 1318 1330
FT REPEAT 1375 AA; 153022 MW; D4B80CBBE0AACE13 CRC64;
SQ SEQUENCE 1375 AA; 153022 MW; D4B80CBBE0AACE13 CRC64;

```



Query Match 92.4%; Score 110; DB 1; Length 1375;  
 Best Local Similarity 95.2%; Pred. No. 4.7e-09;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDTPYLHD 21  
 DB 507 ANDHLSILEAMSYNDTPYLHD 527

RESULT 3  
 GTF2\_STRDO STANDARD; PRT; 1592 AA.  
 AC P27470;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 Streptococcus downei (Streptococcus sobrinus).  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 Streptococcus.  
 NCBI\_TaxID=1317;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=6715;  
 RX MEDLINE=91133227; PubMed=1704006;  
 RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,  
 Kagawa H.;  
 RT "peptide sequences for sucrose splitting and glucan binding within  
 Streptococcus sobrinus glucosyltransferase (water-insoluble glucan  
 synthetase).";  
 RL J. Bacteriol. 173:989-996(1991).  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE  
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
 CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- DISEASE: DENTAL CARIES.  
 CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
 CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES). GTF-S SYNTHESIZES  
 CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
 CC FORMS OF GLUCANS.  
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 CC BINDING PROTEIN FROM S. MUTANS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D90213; BAA14241.1; -;  
 DR PIR: A38175; A38175.  
 DR HSSP: P00695; 2HEE.  
 DR InterPro: IPR002479; CM\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF021473; CM\_binding\_1; 16.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.  
 FT SIGNAL 1 38 POTENTIAL.  
 FT CHAIN 39 1592 GLUCOSYLTRANSFERASE-I.  
 FT DOMAIN 39 1044 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1093 1592 GLUCAN-BINDING (APPROXIMATE).  
 FT REPEAT 1093 1592 6.5 X TANDEM REPEATS.  
 FT REPEAT 1158 1207 2.  
 FT REPEAT 1222 1272 3.

FT REPEAT 1287 1337 4.  
 FT REPEAT 1402 1451 5.  
 FT REPEAT 1514 1563 6.  
 FT REPEAT 1577 1592 7 (INCOMPLETE).  
 SQ SEQUENCE 1592 AA; 176167 MW; BCOA6DD079351ECF CRC64;

Query Match 90.8%; Score 108; DB 1; Length 1592;  
 Best Local Similarity 85.7%; Pred. No. 1.1e-08;  
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDTPYLHD 21  
 DB 477 ANHVSILEAMSYNDTPYLHD 497

RESULT 4  
 GTF1\_STRDO STANDARD; PRT; 1597 AA.  
 AC P11001;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTFI.  
 OS Streptococcus downei (Streptococcus sobrinus).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1317;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MFE28;  
 RX MEDLINE=87308014; PubMed=3040686;  
 RA Ferretti J.J., Gilpin M.L., Russell R.B.;  
 RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus  
 RT sobrinus MFE28";  
 RL J. Bacteriol. 169:4271-4278(1987).  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE  
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
 CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- DISEASE: DENTAL CARIES.  
 CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
 CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES). GTF-S SYNTHESIZES  
 CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
 CC FORMS OF GLUCANS.  
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 CC BINDING PROTEIN FROM S. MUTANS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M17391; AAC63063.1; -;  
 DR InterPro: IPR002479; CM\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF021473; CM\_binding\_1; 19.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.  
 FT SIGNAL 1 38 POTENTIAL.  
 FT CHAIN 39 1597 GLUCOSYLTRANSFERASE-I.  
 FT DOMAIN 39 1050 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1099 1597 GLUCAN-BINDING (APPROXIMATE).  
 FT DOMAIN 1099 1597 1.25 A, 2 B, AND 5 AC REPEATS.  
 FT REPEAT 1099 1132 A REPEAT.

FT	REPEAT	1163	1213	AC REPEAT.
FT	REPEAT	1227	1277	AC REPEAT.
FT	REPEAT	1292	1342	AC REPEAT.
FT	REPEAT	1352	1390	B REPEAT.
FT	REPEAT	1406	1455	AC REPEAT.
FT	REPEAT	1465	1512	B REPEAT.
FT	REPEAT	1519	1568	AC REPEAT.
FT	REPEAT	1582	1597	A REPEAT (INCOMPLETE).
FT	REPEAT	1597	177080	MW: B9E6BA200868798E CRC64
FT	SEQUENCE			

Query Match	90.8%	Score 108;	DB 1;	Length 1597;
Best Local Similarity	85.7%	Pred. No. 1.2e-08;		
Matches 18;	Conservative 3;	Mismatches 0;	Indels 0;	Gaps 0;

```

QY      1 ANDHLSILEAWSNDNTPYLHD 21
        ||:||:||:||:||:||:||||
483 ANNHVSIVEAWSNDNTPYLHD 503

```

```

RESULT      5
GTFS_STRDO
ID          GTFS_STRDO  STANDARD;  PRT;  1365 AA

```

DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE GLUCOSYLTRANSFERASE-S, PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSURASE,  
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).

OS Streptococcus downei (Streptococcus sobrinus).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus  
OX NCBI\_TaxID=1317;

RA. Gilmore K.S., Russell R.R., Ferritt J.J.;  
RT glucosyltransferase that synthesizes soluble glucans.";  
Infect. Immun. 58:2452-2458(1990).

CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DERIS.  
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).

-I- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF  
CC PRIMER GLUCAN UNLIKE GTF-I.  
CC -I- DISEASE: DENTAL CARIES.

CC -1- MISCELLANEOUS: GTF-5 SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA  
CC 1,6-GLUCOSE).  
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN  
CC BINDING PROTEIN FROM S. MUTANS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/sib.ch>). or send an email to [license@sib.ch](mailto:license@sib.ch)).

CC -----  
DR EMBL; M30943; AAA26898.1; -.  
DR PTR; A41483; DR A41483.  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding.1; 10.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries  
FT SIGNAL 1 36 OR 37 (POTENTIAL).  
FT CHAIN 37 1365 GLUCOSYLTRANSFERASE-S.

FT	DOMAIN	37	1050	CATALYTIC (APPROXIMATE).
FT	DOMAIN	1083	1365	GLUCAN-BINDING (APPROXIMATE)
FT	DOMAIN	1083	1365	4.5 X TANDEM REPEATS.
FT	REPEAT	1083	1131	1.
FT	REPEAT	1150	1139	2.
FT	REPEAT	1225	1274	3.
FT	REPEAT	1289	1339	4.
FT	REPEAT	1353	1365	5. (INCOMPLETE).
FT	REPEAT	1365	1365	167296B528EC476 CRC64;
SEQ	SEQUENCE			

Query Match	65.5%	Score	78	DB	1	Length	1365
Best Local Similarity	71.4%	Pred	No.	0.0004			
Matches	15	Conservative	1	Mismatches	5	Indels	0
						Gaps	0

QY 1 ANDHLSILEAWSNDPTPLYHD 21  
| | | | | | | | : |  
Db 467 AIDHLSILEAWSGNDNDYKD 487

RESULT	6	
GTED_STRMU		
ID	GTED_STRMU	STANDARD; PRT; 1462 AA.

DT 01-FEB-1996 (Rel. 33, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE,  
DE (SUCROSE 6-GLUCOSYLTRANSFERASE)).

OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae.  
OC Streptococcus.  
OX NCBI\_TaxID=1309;

RN  
 LI  
 SEQUENCE FROM N.A.  
 RC  
 STRAIN-GS-5;  
 RX  
 MEDLINE-91100958; Pubmed-2148600;  
 RA  
 Honda O., Kato C., Kuramitsu H. K.;  
 RT  
 "Nucleotide sequence of the streptococcus mutans *gtfD* gene encoding  
 the glucosyltransferase-S enzyme.";  
 RL  
 J. Gen. Microbiol. 136:2099-2105(1990).

RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT4239, MT4245, MT4251, MT4467, AND MT6148;  
RX MEDLINE=98231643; Pubmed=5570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
RA Kimura S., Hamada S.;

RT "Molecular analyses of glucosyltransferase genes among strains of  
RT *Streptococcus mutans*,"  
RT FEMS Microbiol. Lett. 161:331-336(1998).  
RL

CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE  
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.

CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
CC CC -1- D-PRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- DISEASE: DENTAL CARIES.  
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
CC CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
CC CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
CC FORMS OF GLUCANS.

CC -I- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
CC BINDING PROTEIN FROM S. MUTANS.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on ways  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M29296: AAA26895.1; -  
DR EMBL: D88653: BAA26103.1; -  
DR EMBL: D88656: BAA26107.1; -  
DR EMBL: D88659: BAA26111.1; -  
DR EMBL: D88662: BAA26115.1; -  
DR EMBL: D89979: BAA26121.1; -  
DR InterPro: IPR002479: CW\_binding.  
DR InterPro: IPR003318: Glyco\_hydro\_70.  
DR Pfam: PF01473: CW\_binding\_1; 11.  
DR Pfam: PF03324: Glyco\_hydro\_70; 1.  
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.  
FT SIGNAL 1  
FT CHAIN ? 1462  
FT DOMAIN 1332 1423  
FT REPEAT 1232 1295  
FT REPEAT 1296 1359  
FT REPEAT 1360 1423  
FT REPEAT 58 58  
FT VARIANT 68 68  
FT VARIANT 81 81  
FT VARIANT 113 113  
FT VARIANT 122 122  
FT VARIANT 132 132  
FT VARIANT 135 135  
FT VARIANT 202 202  
FT VARIANT 255 255  
FT VARIANT 275 275  
FT VARIANT 288 288  
FT VARIANT 301 301  
FT VARIANT 313 313  
FT VARIANT 317 317  
FT VARIANT 328 328  
FT VARIANT 350 350  
FT VARIANT 628 628  
FT VARIANT 688 688  
FT VARIANT 726 726  
FT VARIANT 730 730  
FT VARIANT 964 964  
FT VARIANT 1019 1019  
FT VARIANT 1059 1060  
FT VARIANT 1060 1060  
FT VARIANT 1080 1080  
FT VARIANT 1142 1142  
FT VARIANT 1198 1198  
FT VARIANT 1220 1220  
FT VARIANT 1280 1280  
FT VARIANT 1282 1282  
FT VARIANT 1290 1290  
FT VARIANT 1311 1311  
FT VARIANT 1403 1403  
FT VARIANT 1425 1425  
FT VARIANT 1449 1449  
FT CONFLICT 1428 1462  
FT SEQUENCE 1462 AA: 163512 MW: 5C6541F0DCB0DF00 CRC64;  
Query Match 63.0%; Score 75; DB 1; Length 1462;  
Best Local Similarity 71.4%; Pred. No. 0.0014;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
QY 1 ANDHLSLEAWSNDNDPYLHD 21  
I :||||| I I

DB 495 AINHLSLEAWSNDNDPYXND 515  
RESULT 7  
G3PC\_LEIME STANDARD; PRT; 330 AA.  
AC 001558;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, CYTOSOLIC (EC 1.2.1.12)  
GN (GAPDH).  
DE (GAPDH).  
OS Leishmania mexicana.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
OX NCBI\_TaxID=5665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SSP. MEXICANA;  
RX MEDLINE=93063042; PubMed=1435864;  
RA Hannafert V., Blaauw M., Kohl L., Allert S., Oppendoes F.R.,  
RA Michels P.A.M.;  
RT "Molecular analysis of the cytosolic and glycosomal glyceraldehyde-3-  
RT phosphate dehydrogenase in Leishmania mexicana.";  
RL Mol. Biochem. Parasitol. 55:115-126(1992).  
CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE  
CC + NAD(+) = 1,3-DIPHOSPHATEGLYCERATE + NADH.  
CC -1- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.  
CC -1- SUBUNIT: HOMOTETRAMER.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE  
CC DEHYDROGENASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X65220: CAA46323.1; -  
DR PIR: S25142; S25142.  
DR PIR: B48445; B48445.  
DR HSSP: P06977; 1GAD.  
DR InterPro: IPR000173: GAP\_DH.  
DR Pfam: PF00044; gpdh; 1.  
DR PRINTS: PR00078: G3PDHGNASE.  
DR PROSITE: PS00071; GAPDH; 1.  
KW Glycolysis; Oxidoreductase; NAD.  
FT INIT\_MET 0  
FT BINDING 148 148  
FT ACT\_SITE 175 175  
FT SEQUENCE 330 AA: 35511 MW: ED4B6D8BEA207F1E CRC64;  
Query Match 47.9%; Score 57; DB 1; Length 330;  
Best Local Similarity 52.6%; Pred. No. 0.15;  
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;  
QY 2 NDHLSLEAWSNDNDPYLHD 20  
DB 300 NDHFVKLVSWYDNETGYSH 318  
RESULT 8  
ID U16\_EBV STANDARD; PRT; 336 AA.  
AC P03221;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE BGLF2 PROTEIN.

```

CN      BGHF2.
OS      Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC      Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC      Gammaherpesvirinae; Lymphocryptovirus.
OX      NCBI_TaxID=103377;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=84270667; PubMed=6087149;
RA      Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA      Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA      Tuffnell P.S., Barrett B.G.;
RT      "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL      Nature 310:207-211(1984).
RP      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=92113548; PubMed=1662696;
RA      Chen M.-R., Hsu T.-Y., Lin S.-W., Chen J.-Y., Yang C.-S.;
RT      "Cloning and characterization of cDNA clones corresponding to
RT      transcripts from the BamHI G region of the Epstein-Barr virus genome
RT      and expression of BGHF2.";
RL      J. Gen. Virol. 72:3047-3055(1991).
CC      -I- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UN16,
CC      HSV-6 ORF1R, EBV-1 46, HCMV UL94, EBV BDLF2, AND VZV 44.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL; V01514; AAA45871.1; -
DR      EMBL; V01555; CAA24831.1; -
DR      EMBL; S77132; AAB21113.1; -
DR      PIR; A03784; QOBE40.
DR      PIR; S33036; S33036.
RW      Late protein.
SQ      SEQUENCE 336 AA: 36888 MW: 8409374416D5584C CRC64:

```

```

Query Match          41.2%   Score 49; DB 1; Length 336;
Best Local Similarity 41.2%   Pred. No. 2.7;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY      1  ANDHLSTLEAMSNDNTP 17
      | 1::11 1::11
      216 AGAHVNIILRGWTEDDSP 232

RESULT 9
G3PC_TRYBB
ID      G3PC_TRYBB      STANDARD;      PRT;      330 AA.
AC      P10097;
DT      01-MAR-1989 (Rel. 10, Created)
DT      01-MAY-1992 (Rel. 22, Last sequence update)
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, CYTOSOLIC (EC 1.2.1.12)
DE      (GAPDH).
OS      Trypanosoma brucei brucei.
OS      Eukaryota, Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX      NCBI_TaxID=5702;
RN      (1)
RN      SEQUENCE FROM N.A.
RP      STRAIN=427;
RX      MEDLINE=91249838; PubMed=2040303;
RA      Michels P.A.M., Marchand M., Kohl L., Allert S., Wierenga R.K.,
RA      Opperdoes F.R.;
RT      "The cytosolic and glycosomal isoenzymes of glyceraldehyde-3-phosphate
RT      dehydrogenase in Trypanosoma brucei have a distant evolutionary
RT      relationship."
RL      Eur. J. Biochem. 198;421-428(1991).
RN      (2)

```

```

RP PRELIMINARY SEQUENCE OF 1-85
RX MEDLINE=87161817; PubMed=3830153;
RA Misset O., van Beunten J., Lambelr A.M., van der Meer R.,
RA Operdides F.R.;
RA "glyceraldehyde-phosphate dehydrogenase from Trypanosoma brucei.
RT Comparison of the glycosomal and cytosolic isoenzymes.";
RL Eur. J. Biochem. 162:501-507(1987).
CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE
CC + NAD(+) -> 1,3-DIPHOSPHATEGLICERATE + NADH.
CC -1- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
CC DEHYDROGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, X53472; CA937568.1; -.
DR PIR, S16091; DEUTGC.
DR HSSP, P06977; IGAE.
DR InterPro, IPR000173; GAP_DH.
DR Pfam, PF00044; gpdb. 1.
DR PRINTS, PR00078; G3PDHDEGNASE.
DR PROSITE, PS00071; GAPDH; 1.
KW Glycolysis; Oxidoreductase; NAD.
FT INIT_MET 0
FT BINDING 149 149 GLYCERALDEHYDE 3-PHOSPHATE.
FT ACT_SITE 176 176 ACTIVATES THIOL GROUP DURING CATALYSIS.
FT VARIANT 72 72 R -> I.
FT CONFLICT 56 56 A -> T (IN REF. 2).
FT CONFLICT 60 60 Q -> K (IN REF. 2).
SQ SEQUENCE 330 AA; 35503 MW; 7E1F0AFD091E75 CRC64;

```

```

Query Match          39.9%; Score 47.5; DB 1; Length 330;
Best Local Similarity 43.5%; Pred. No. 4.5;
Matches 10; Conservative 4; Mismatches 6; Indels 3; Gaps 1;

QY      2 NDHLSLEAWSNDNPPY--LMD 21
      11: 1 1 1 1 1 1 1 1 1 1
DB      301 NDNFVKLVSWYDNERTGYSNKHVD 323

RESULT 10
YG33_YEAST
AC      YG33_YEAST STANDARD; PRT; 275 AA.
DT      P53271;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DE      01-NOV-1997 (Rel. 35, Last annotation update)
DE      HYPOTHEETICAL 31.8 KDA PROTEIN IN NUP57-MEPI INTERGENIC REGION.
GN      YGR120C OR G6324.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX      NCBI_TaxID=4932;
RN
RP      SEQUENCE FROM N.A.
RC      STRAIN=S288C / FY1679;
RX      MEDLINE=97197982; PubMed=9046098;
RA      van Dyck L., Tetzelin H., Purnelle B., Goffeau A.;
RT      "An 18.3 kb DNA fragment from yeast chromosome VII carries four
RT      unknown open reading frames, the gene for an Asn synthase, remnants
RT      of Ty and three tRNA genes."
RL      Yeast 13:171-176(1997).
CC      -----
CC      This SWISS-PROT entry is copyright  It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X83099; CAA58155.1; -  
DR EMBL: Z72905; CAA97130.1; -  
DR SGD: S0003352; SEC35.  
KW Hypothetical protein.  
SQ SEQUENCE 275 AA; 31799 MW; AA102D086FE3FAD7 CRC64;

Query Match 38.7%; Score 46; DB 1; Length 275;  
Best Local Similarity 50.0%; Pred. No. 6.1;  
Matches 7; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 7 ILEAMSDNTPYLH 20  
|||:|:|:|:|:|:|  
217 ILESCADSNSPRIH 230

RESULT 11  
G3P\_CHLPN STANDARD; PRT: 335 AA.  
ID G3P\_CHLPN 09J0H7;  
AC 09J7T0: 09J0H7;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.12) (GAPDH).  
GN GAP OR GAPA OR CPN0624 OR CP0123.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83538;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CML029;  
RX MEDLINE=99206606; PubMed=10192388;  
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
RT Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
RL "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
RN Nat. Genet. 21:385-389(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AR39;  
RX MEDLINE=20150255; PubMed=10684935;  
RA White T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RT Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
RL Eisen J., Fraser C.M.;  
RN "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
RT pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-J138;  
RX MEDLINE=20330349; PubMed=10871362;  
RA Shiba T., Ishikawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
RT Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
RN "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
RT from Japan and CML029 from USA";  
RL Nucleic Acids Res. 28:2311-2314(2000).  
CC -I- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE  
CC + NAD(+) = 1,3-DIPHOSPHATE + NADH.  
CC -I- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.  
CC -I- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
CC -I- SUBCELLULAR LOCATION: CYTOSOL.  
CC -I- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE  
CC DEHYDROGENASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AE001647; AAD18763.1; -  
DR EMBL: AE002173; AAF38006.1; -  
DR EMBL: AP002547; BAA9831.1; -  
DR HSSP: P06977; 1GAE.  
DR TIGR: CP0123; -  
DR InterPro: IPR000173; GAP\_DH.  
DR Pfam: PF00044; gpdb.1.  
DR PRINTS: PR00078; G3PDHGRNASE.  
DR PROSITE: PS00071; GAPDH.1.  
KW Glycolysis; Oxidoreductase; NAD; Complete proteome.  
FT BINDING 151  
FT ACT\_SITE 178 178  
FT SEQUENCE 335 AA; 36837 MW; C86DDE3AD3ADFEE CRC64;

Query Match 38.7%; Score 46; DB 1; Length 335;  
Best Local Similarity 52.9%; Pred. No. 7.7;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 2 NDHLSILEAMSDNTPY 18  
|||:|:|:|:|:|:|  
DB 303 NDRFKLVAMWDNENGY 319

RESULT 12  
PCHA\_PSEAE STANDARD; PRT: 476 AA.  
ID PCHA\_PSEAE  
AC 051508;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE SALICYLATE BIOSYNTHESIS ISOCORISMATE SYNTHASE (EC 5.4.99.6).  
GN PCHA OR PA4231.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 15692 / PA01;  
RX MEDLINE=2043737; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RT Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RL Gardner R.L., Golltry L., Tolentino E., Westbrook-Medman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RT Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RN "Structural genes for salicylate biosynthesis from chorismate in  
RT Pseudomonas aeruginosa.";  
RL Mol. Gen. Genet. 249:217-228(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 15692 / PA01;  
RX MEDLINE=2043737; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RT Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RL Gardner R.L., Golltry L., Tolentino E., Westbrook-Medman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RT Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RN "Structural genes for salicylate biosynthesis from chorismate in  
RT Pseudomonas aeruginosa.";  
RL Nature 406:959-964(2000).  
CC -I- FUNCTION: INVOLVED IN THE CONVERSION OF CHORISMATE TO SALICYLATE  
CC (PROBABLE).  
CC -I- CATALYTIC ACTIVITY: CHORISMATE = ISOCORISMATE.  
CC -I- PATHWAY: SALICYLATE BIOSYNTHESIS.  
CC -I- SIMILARITY: STRONG, TO OTHER ISOCORISMATE SYNTHASES; WEAK, TO  
CC TRE AND PAB.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X82644; CAA57969.1; -  
 DR EMBL: AE04840; AAG07619.1; -  
 DR InterPro: IPR000350; Choriismate\_bind.  
 DR Pfam: PF00425; choriismate\_bind.1.  
 DR Prodom: PD00779; Choriismate\_bind.1.  
 DR Isomerase; Complete proteome.  
 SQ SEQUENCE 476 AA; 52071 MW; D0F6AFD9C9CF9CFS CRC64;  
 Query Match 38.7%; Score 46; DB 1; Length 476;  
 Best Local Similarity 44.4%; Pred. No. 12;  
 Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;  
 Oy 3 DRLSLEANSNDPYLH 20  
 Db 44 DPLQVFGAMDDROTPLY 61  
 RESULT 13  
 GALT\_HUMAN STANDARD; PRT; 236 AA.  
 ID GALT\_HUMAN  
 AC 014353;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE GUANIDINACETATE N-METHYLTRANSFERASE (EC 2.1.1.2).  
 GN GALT.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=96138544; PubMed=8547310;  
 RA Isbrandt D., von Figura K.;  
 RT "Cloning and sequence analysis of human guanidinoacetate N-  
 methyltransferase cDNA.";  
 RT Biochim. Biophys. Acta 1264:265-267 (1995).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97472276; PubMed=9325156;  
 RA Jenne D.E., Olsen A.S., Zimmer M.;  
 RT "The human guanidinoacetate methyltransferase (GALT) gene maps to a  
 RT syntenic region on 19p13.3, homologous to band C of mouse chromosome  
 RT 10, but GALT is not mutated in jittery mice.";  
 RT Biochem. Biophys. Res. Commun. 238:723-727 (1997).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RA Isbrandt D., Schmidt A.;  
 RT "Gene structure of human guanidinoacetate N-methyltransferase.";  
 RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 [4]  
 RP SEQUENCE FROM N.A.  
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,  
 RA Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,  
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,  
 RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,  
 RA Liu S., Altix C., Andreise T., Frankheim M., Amico-Keller G.,  
 RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,  
 RA Kromboller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,  
 RA Kobayashi A., Olsen A.S., Carrano A.V.;  
 RA Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + GUANIDOACETATE =

-----  
 CC S-ADENOSYL-L-HOMOCYSTEINE + CREATINE.  
 CC -1- PATHWAY: LAST STEP OF CREATINE BIOSYNTHESIS.  
 CC -1- DISEASE: DEFECTS IN GALT ARE RESPONSIBLE FOR NEUROLOGIC SYNDROMES  
 CC AND MUSCULAR HYPOTONIA. BIOCHEMICALLY IT IS CHARACTERIZED BY LOW  
 CC EXCRETION OF CREATINE, DEFICIENCY OF CREATINE AND CREATINE  
 CC PHOSPHATE, AND SIMULTANEOUS ACCUMULATION OF GUANIDINOACETATE IN  
 CC BRAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z49878; CAA90035.1; -  
 DR EMBL: AF010248; AAD04781.1; -  
 DR EMBL: AF010246; AAD04781.1; JOINED.  
 DR EMBL: AF010247; AAD04781.1; JOINED.  
 DR EMBL: AF188893; AAF01461.1; -  
 DR EMBL: AC005329; AAC27668.1; -  
 DR MIM: 601240; -  
 DR Transferase; Methyltransferase.  
 SQ SEQUENCE 236 AA; 26318 MW; 6B8E845CE56189F5 CRC64;  
 Query Match 37.8%; Score 45; DB 1; Length 236;  
 Best Local Similarity 44.0%; Pred. No. 7.3;  
 Matches 11; Conservative 3; Mismatches 3; Indels 8; Gaps 2;  
 Oy 1 ANDHSIL-----EAMSDNDPYLH 20  
 Db 30 ADPLRLILKPKVEMK---ETPYMH 51  
 RESULT 14  
 GUNF\_FUSOX STANDARD; PRT; 385 AA.  
 ID GUNF\_FUSOX  
 AC P46239;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE PUTATIVE ENDOGLUCANASE TYPE F PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-  
 DE GLUCANASE) (CELLULOSE).  
 OS Fusarium oxysporum.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreales; mitosporic Hypocreales; Fusarium.  
 OX NCBI\_TaxID=5507;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95047531; PubMed=7959045;  
 RA Sheppard P.O., Grant F.J., Oort F.J., Sprecher C.A., Foster D.C.,  
 RA Hagen F.S., Upshall A., McKnight G.L., O'Hara P.J.;  
 RT "The use of conserved cellulase family-specific sequences to clone  
 RT cellulase homologue cDNAs from Fusarium oxysporum.";  
 RT Gene 150:163-167 (1994).  
 RL -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC  
 CC LINKAGES IN CELLULOSE.  
 CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY F (FAMILY 10 OF GLYCOSYL  
 CC HYDROLASES).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L29380; AAA65588.1; -  
 DR HSP; P56588; IBG4.

DR	InterPro: IPR000254; CBD_fungrl.
DR	InterPro: IPR001000; Glyco_hydr0_10.
DR	Pfam: PF00734; CBD_1; 1.
DR	Pfam: PF00331; Glyco_hydr0_10; 1.
DR	PRINTS: PR00134; GLYXDRASE10.
DR	ProDom: PD001821; CBD_fungral; 1.
DR	SMART: SMO0236; FCBP; 1.
DR	ProSITE: PS00562; CBD_FUNGAL; 1.
DR	ProSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.
KW	Cellulose degradation; Hydrolase; signal.
FT	SIGNAL 1 19 POTENTIAL.
FT	CHAIN 20 385 PUTATIVE ENDOGLUCANASE TYPE F.
FT	DOMAIN 25 52 CELLULOSE-BINDING.
FT	DOMAIN 53 84 LINKER.
FT	DOMAIN 85 385 CATALYTIC.
FT	ACT_SITE 209 209 PROTON DONOR (BY SIMILARITY).
FT	ACT_SITE 321 321 NUCLEOPHILE (BY SIMILARITY).
SO	SEQUENCE 385 AA; 41225 MW; B3C3807C07D3C0EC CR64;

Query Match	37.8%	Score 45	DB 1	Length 385
Best Local Similarity	24.4%	Pred. NO. 13		
Matches 10; Conservative	5	Mismatches	6	Indels 20; Gaps 1;

QY 1 ANDHLSILEA-----WSDNDTPYLHD 21  
|||::| | | | |  
Db 329 ANDYLTVMNACLAVPKCVGITVWGVSDKDSWRPGDNPPLYD 369

RESULT	15
AMP2_STRCO	
ID	AMP2_STRCO
AC	060394;
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	XAA-PRO AMINOPEPTIDASE II (EC 3.4.11.9) (X-PRO AMINOPEPTIDASE II)
DE	(AMINOPEPTIDASE P II) (APP) (PEPP II) (AMINOACYLPROLINE AMINOPEPTIDASE II).
GN	PEPP2 OR SCG61.34C.
OS	Streptomyces coelicolor, and Streptomyces lividans.
OC	Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae;
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX	NCBI_Taxid:1902, 1916;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	SPECIES=S.coelicolor; STRAIN=A3(2);
RA	Oliver K., Harris D., Bentley S.D., Parkhill J., Barrell B.G.,
RA	Rajandream M.A.; Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

SEQUENCE FROM N.A.  
RC SPECIES-S: LIVIDANS; STRAIN-66;  
RX MEDLINE=94176084; PubMed=7765336;  
RA Butler M.J., Aphale J.S., Dizonno M.A., Krysman P., Walczyk E.,  
RA Malek L.T.;  
RT "Intracellular aminopeptidases in Streptomyces lividans 66.";  
RL J. Intl. Microbiol. 13:24-29(1994).  
CC -1- CATALYTIC ACTIVITY: RELEASE OF ANY N-TERMINAL AMINO ACID,  
CC INCLUDING PROLINE, THAT IS LINKED WITH PROLINE, EVEN FROM A  
CC DIPEPTIDE OR TRIPEPTIDE.  
CC -1- COFACTOR: ZINC OR MANGANESE.  
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -1- MISCELLANEOUS: IN STREPTOMYCES LIVIDANS, THERE ARE TWO GENES  
CC (PEPPI AND PEP2P) WHICH ENCODE AMINOPEPTIDASE P.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24B.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).

```
CC or send an email to license@sb-sib.ch.
CC -----
DR EMBL; AL359949; CAB95809.1; -.
DR EMBL; U23174; AAB00325.1; -.
DR MEROPS; M24.025; -.
DR InterPro; IPR000994; Peptidase_M24.
DR InterPro; IPR001131; Pro_dipeptidase.
DR Pfam; PF00557; Peptidase_M24; 1.
DR PROSITE; PS00491; PROLINE_PEPTIDASE; 1.
KW Hydrolase; Aminopeptidase; Manganese; Zinc.
SQ SEQUENCE 470 AA; 51924 MW; 0D3D33CE2565236C2 CRC64;
```

Query Match	37.4%	Score 44.5	DB 1	Length 470
Best Local Similarity	47.1%	Pred. No. 20		
Matches	8	Conservative	4	Mismatches 4; Indels 1; Gaps 1;
QY	1	ANDHSLIENASNDTP	17	
		:   :   :   :   :		
Db	262	AGEHATLTH-WTDNDGP	277	

```
Search completed: March 27, 2002, 14:27:02
Job time: 1644 sec
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:26:03 ; Search time 188.53 Seconds

(without alignments)  
16.293 Million cell updates/sec

Title: US-09-290-049a-10  
Perfect score: 119  
Sequence: 1 ANDHLSLEAWSNDPTYLHD 21

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: SPREMBL\_17:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mhc:\*  
8: sp\_mammal:\*  
9: sp\_organelle:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

ult No.	Score	Query Match	Length	ID	Description
1	119	100.0	1390	2	069385 streptococc
2	119	100.0	1455	2	069391 streptococc
3	110	92.4	1455	2	069382 streptococc
4	110	92.4	1455	2	069388 streptococc
5	110	92.4	1455	2	069397 streptococc
6	108	90.8	1590	2	055263 streptococc
7	108	90.8	1590	2	059983 streptococc
8	76	65.5	1338	2	09WXJ4 streptococc
9	76	63.9	1016	2	09LCJ7 streptococc
10	74	62.2	1477	2	091466 leucostoc
11	74	62.2	1508	2	052224 leucostoc
12	74	62.2	1508	2	09EZH5 leucostoc
13	73	61.3	1575	2	091CH3 streptococc
14	73	61.3	1577	2	054178 streptococc
15	71	59.7	1527	2	09ZAR4 streptococc
16	66	55.5	1449	2	068542 streptococc
17	66	55.5	1449	2	055264 streptococc
18	63	52.9	1577	2	055265 streptococc
19	62	52.1	1512	2	09WXJ5 streptococc

20	60	50.4	2057	2	09RE05	09e05 leucostoc
21	58	48.7	1518	2	000600	000600 streptococc
22	57	47.9	1599	2	000599	000599 streptococc
23	52	43.7	4848	2	007944	007944 streptomyc
24	51	42.9	435	10	091U68	091u68 arabidopsis
25	50	42.0	549	10	091U92	091u92 arabidopsis
26	49.5	41.6	549	10	043274	043274 zea mays (m
27	49	41.2	601	2	005566	005566 mycobacteri
28	48	40.3	490	2	006579	006579 mycobacteri
29	47.5	39.9	519	10	09ZUB6	09zub6 arabidopsis
30	47.5	39.9	549	10	091LR2	091lr2 oryza sativ
31	47.5	39.9	549	10	09FRX7	09frx7 oryza sativ
32	47	39.5	175	10	09LMJ8	09lmj8 arabidopsis
33	47	39.5	378	10	09SZH3	09szh3 arabidopsis
34	47	39.5	524	2	09KS18	09ks18 vibrio chol
35	47	39.5	525	3	056679	056679 schizosacch
36	46	38.7	195	13	09DFN6	09dfn6 gliliichthys
37	46	38.7	2567	5	09U0J6	09u0j6 plasmodium
38	45.5	38.2	96	2	09JST4	09jst4 neisseria m
39	45	37.8	286	2	09RDG2	09rdg2 streptomyc
40	45	37.8	333	13	09PTW5	09ptw5 paratichthy
41	45	37.8	947	10	023136	023136 arabidopsis
42	45	37.8	1159	4	09UE39	09ue39 homo sapien
43	44.5	37.4	210	2	0910Y4	0910y4 pseudomonas
44	44.5	37.4	499	5	09VJ56	09vj56 drosophila
45	44	37.0	44	4	016768	016768 homo sapien

## ALIGNMENTS

## RESULT 1

ID 069385 PRELIMINARY; PRT: 1390 AA.  
AC 069385;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-ST.  
GN GTFC.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=109;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT4245;  
RX MEDLINE=98271643; PubMed=9570124;  
RA Fujiwara T., Iferao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
RA Kimura S., Yamada S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
RT Streptococcus mutans";  
RL EMBL: D88655; BAA26106.1; ;  
DR EMBL; D88655; BAA26106.1; ;  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 7.  
DR Pfam: PF03324; Glyco\_hydro\_70; 1.  
KW transferase.  
SQ SEQUENCE 1390 AA; 155375 MW; 8847EA956EF05E9F CRC64;

Query Match 100.0%; Score 119; DB 2; Length 1390;  
Best Local Similarity 100.0%; Pred. No. 7.5e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSLEAWSNDPTYLHD 21  
|||||  
Db 507 ANDHLSLEAWSNDPTYLHD 527  
RESULT 2  
069391

ID 069391 PRELIMINARY; PRT; 1455 AA.  
 AC 069391:  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
 DE GLUCOSYLTRANSFERASE-SI.  
 GN GTFc.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT4251;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Teroo Y., Hoshino T., Kawabata S., Sobue S.,  
 Kimura S., Hamada S.;  
 "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans."  
 RL EMBL; D88661; BAA26114.1; -.  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 9.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 SQ SEQUENCE 1455 AA; 162804 MW; 683A359D873E9E1A CRC64;

Query Match 100.0%; Score 119; DB 2; Length 1455;  
 Best Local Similarity 100.0%; Pred. No. 7.9e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHLSILEAMSDNDPTPLHD 21  
 DB 507 ANDHLSILEAMSDNDPTPLHD 527

RESULT 3  
 ID 069382 PRELIMINARY; PRT; 1455 AA.  
 AC 069382:  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
 DE GLUCOSYLTRANSFERASE-SI.  
 GN GTFc.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT8148;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Teroo Y., Hoshino T., Kawabata S., Sobue S.,  
 Kimura S., Hamada S.;  
 "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans."  
 RL EMBL; D88652; BAA26102.1; -.  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 9.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 SQ SEQUENCE 1455 AA; 162969 MW; 27DAD3A1ECCA2939 CRC64;

Query Match 92.4%; Score 110; DB 2; Length 1455;  
 Best Local Similarity 95.2%; Pred. No. 2e-08;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ANDHLSILEAMSDNDPTPLHD 21  
 DB 507 ANDHLSILEAMSDNDPTPLHD 527

RESULT 4  
 ID 069388 PRELIMINARY; PRT; 1455 AA.  
 AC 069388:  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
 DE GLUCOSYLTRANSFERASE-SI.  
 GN GTFc.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT4239;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Teroo Y., Hoshino T., Kawabata S., Sobue S.,  
 Kimura S., Hamada S.;  
 "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans."  
 RL EMBL; D88658; BAA26110.1; -.  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 10.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 SQ SEQUENCE 1455 AA; 163045 MW; 6D90A4978D35DD82 CRC64;

Query Match 92.4%; Score 110; DB 2; Length 1455;  
 Best Local Similarity 95.2%; Pred. No. 2e-08;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ANDHLSILEAMSDNDPTPLHD 21  
 DB 507 ANDHLSILEAMSDNDPTPLHD 527

RESULT 5  
 ID 069397 PRELIMINARY; PRT; 1455 AA.  
 AC 069397:  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
 DE GLUCOSYLTRANSFERASE-SI.  
 GN GTFc.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT4467;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Teroo Y., Hoshino T., Kawabata S., Sobue S.,  
 Kimura S., Hamada S.;  
 "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans."  
 RL EMBL; D88978; BAA26120.1; -.  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 9.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.



RX MEDLINE=20169623; PubMed=10705445;  
RA Funane K., Mizuno K., Takahara H., Kobayashi M.;  
RT "Gene encoding a dextrinase-1-like protein in *Leuconostoc*  
RT *mesenteroides* NRRL B-512F.";  
RL Biosci. Biotechnol. Biochem. 64:29-38(2000).  
DR EMBL, AB020020; AAA00527.1.;  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF03282; Glyco\_hydro\_70; I.  
SQ SEQUENCE 1016 AA; 110343 MW; 8896EFDE13CCBA7 CRC64;

Query Match	63.9%;	Score 76;	DB 2;	Length 1016;
Best Local Similarity	61.9%;	Pred. NO. 0.0025;		
Matches 13; Conservative	3;	Mismatches 5;	Indels 0;	Gaps 0;

```
QY      1 ANDHLSILEAWSNDNDTPYLDH 21
          :|:|:|:|:|:|:|:|:|:|
548      SNQHVSILEDWSDNDAEYVKD 568
```

RESULT	10	
Q9L466		
ID	PRELIMINARY;	PRT; 1477 AA
Q9L466		
Q9L466		

pfam: PF02324; Glyco\_hydro\_70; 1. Transferase; Glycosyltransferase. SEQUENCE 1477 AA; 164886 MW; E6F5710DEDFCB831 CRC64;

```
QY 1 ANDHLSLEAWSNDNDTPYLHD 21
    || ||||| || | : |
Db 532 ANQHLSTLEDWSHNDPEYVKD 552
```

RESULT	11	
052224		
ID	052224	PRELIMINARY;
052224		PRT; 1508 AA

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NRRL B-1299;  
 RA Monchois V., Renaud-Slineau M., Monsan P., Willemot R.M.;  
 RL FEMS Microbiol. Lett. 0:0-0(1998).  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) -> D-  
 CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).  
 CC EMBL: AF030129; AAB95453.1; -.  
 DR InterPro: IPR002479; CW binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 14.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KM Transferrase: Glycosyltransferase.  
 SQ SEQUENCE 1508 AA; 168511 MW; E70CCECB5A70D1F0 CRC64; .

Query Match	62.2%	Score 74;	DB 2;	Length 1508;
Best Local Similarity	66.7%	Pred. No. 0.0081;		
Matches 14; Conservative	1;	Mismatches 6;	Indels 0;	Gaps 0;

Qy 1 ANDHLSLEAWSNDNDTPYLHD 21  
Db 563 ANQHLSTLEWDWSHNDPEYVKD 583

RESULT	12	
Q9EZH5		
ID	Q9EZH5	PRELIMINARY;
AC	CONFIRMED	PRT; 1508 AA

Query Match	62.2%	Score 74;	DB 2;	Length 1508;
Best Local Similarity	66.7%	Pred. No. 0.0081;		
Matches 14; Conservative	1;	Mismatches 6;	Indels 0;	Gaps 0;

```
QY      1 ANDHLSILEAWSDNDTPYLHD 21
          || ||||| || | : |
Db      563 ANQHLSILEDWSDNDPEYKD 583
```

RESULT	13
09LCH3	
ID	09LCH3
AC	09LCH3; PRELIMINARY; PRT; 1575 AA.
DT	01-OCT-2000 (Tremblrel, 15, Created)
DT	01-OCT-2000 (Tremblrel, 15, last sequence update)
DT	01-JUN-2001 (Tremblrel, 17, last annotation update)
DE	GLUCOSYLTRANSFERASE.
GN	GTFR.
OS	Streptococcus oralis.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.

OX NCBI\_TaxID=1303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC10557;  
 RX MEDLINE=20231779; PubMed=10768934;  
 RA Fujimura T., Hoshino T., Ooshima T., Sobue S., Hamada S.;  
 RT "Purification, characterization, and molecular analysis of the gene  
 encoding glucosyltransferase from *Streptococcus oralis*.";  
 RL Infect. Immun. 68:2475-2483(2000).  
 DR EMBL: AB025228; BAA95201.1; -;  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 17.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 SQ SEQUENCE 1575 AA; 176792 MW; 772A26E4D7CE543 CRC64;

Query Match 61.3%; Score 73; DB 2; Length 1575;  
 Best Local Similarity 77.8%; Pred. No. 0.012;  
 Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 4 HLSTLEAWSNDNDPYLHD 21  
 |||||||||  
 549 HLSTLEAWSNDNDPYNKD 566

RESULT 14  
 054178  
 ID 054178 PRELIMINARY; PRT; 1577 AA.  
 AC 054178; 054247;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE.  
 GN GTFG.  
 OS *Streptococcus gordonii* Challis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 CC Streptococcus.  
 OX NCBI\_TaxID=29390;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CHALLIS;  
 RX MEDLINE=96157084; PubMed=8586195;  
 RA Vickerman M.M., Sulavik M.C., Clewell D.B.;  
 RT "Molecular analysis of *Streptococcus gordonii* glucosyltransferase  
 phase variants.";  
 RL Dev. Biol. Stand. 85:309-314(1995).  
 [2]  
 SEQUENCE OF 1-96 FROM N.A.  
 STRAIN-CHALLIS;  
 MEDLINE=92276337; PubMed=1534326;  
 RA Sulavik M.C., Tardif G., Clewell D.B.;  
 RT "Identification of a gene, rgg, which regulates expression of  
 glucosyltransferase and influences the Spp phenotype of *Streptococcus*  
*gordonii* Challis.";  
 RL J. Bacteriol. 174:3577-3586(1992).  
 DR EMBL: U12643; AAC43483.1; -;  
 DR EMBL: M89776; AAC26969.1; -;  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 18.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 SQ SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5E08D18 CRC64;

Query Match 61.3%; Score 73; DB 2; Length 1577;  
 Best Local Similarity 77.8%; Pred. No. 0.012;  
 Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 HLSTLEAWSNDNDPYLHD 21  
 |||||||||

Db 551 HLSTLEAWSNDNDPYNKD 568

RESULT 15  
 09ZAR4  
 ID 09ZAR4 PRELIMINARY; PRT; 1527 AA.  
 AC 09ZAR4;  
 DT 01-MAY-1999 (TReMBLrel. 10, Created)  
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE.  
 GN DEX.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 CC Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NRRL B-512-F;  
 RA Bhatnagar R., Singh D.K.S.;  
 RT "Cloning and Molecular Characterization of Dextranucrase Gene from  
*Leuconostoc mesenteroides* NRRL B-512F.";  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U81374; AAD10952.1; -;  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 16.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1527 AA; 169709 MW; 1DFAF237C743398 CRC64;

Query Match 59.7%; Score 71; DB 2; Length 1527;  
 Best Local Similarity 66.7%; Pred. No. 0.024;  
 Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANDHLSTLEAWSNDNDPYLHD 21  
 |||||||||  
 Db 581 ANQHLSTLEAWSNDNDPYLHD 601

Search completed: March 27, 2002, 14:26:04  
 Job time: 1676 sec





```
RESULT 2
US-09-740-274-2
; Sequence 2, Application US/09740274
; GENERAL INFORMATION:
; APPLICANT: NICHOLS, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2
```

```
Query Match          100.0%; Score 95; DB 21; Length 1475;
Best Local Similarity 100.0%; Pred. No. 4.5e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 VPSYSFIRAHDSFVODLI 19  
|||||

Db 552 VPSYSFIRAHDSFVODLIA 570

```
RESULT 3
US-09-290-049-15
; Sequence 15, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: FDC98-01P2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 22
; TYPE: PRT
; ORGANISM: S. mutans
US-09-290-049-15
```

```
Query Match          95.8%; Score 91; DB 16; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.3e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 VPSYSFIRAHDSFVODLI 18  
|||||

Db 1 VPSYSFIRAHDSFVODLI 18

```
RESULT 4
US-09-740-274-4
; Sequence 4, Application US/09740274
; GENERAL INFORMATION:
; APPLICANT: NICHOLS, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4
```

```
Query Match          95.8%; Score 91; DB 21; Length 1375;
Best Local Similarity 100.0%; Pred. No. 2.1e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 VPSYSFIRAHDSFVODLI 18  
|||||

Db 578 VPSYSFIRAHDSFVODLI 595

```
RESULT 5
US-09-290-049-17
; Sequence 17, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: FDC98-01P2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 22
; TYPE: PRT
; ORGANISM: S. downei
US-09-290-049-17
```

```
Query Match          90.5%; Score 86; DB 16; Length 22;
Best Local Similarity 94.4%; Pred. No. 1e-07;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

Qy 1 VPSYSFIRAHDSFVODLI 18  
|||||

Db 1 VPSYSFIRAHDSFVODLI 18



RESULT 6  
US-09-290-049-2  
; Sequence 2, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: FDC98-01P2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; EARLIER FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: HDS peptide  
US-09-290-049-2

Query Match 88.9%; Score 84.5; DB 16; Length 20;  
Best Local Similarity 95.0%; Pred. No. 1.6e-07;  
Matches 19; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 VPSYSFIRAHDSVODLIA 19  
|||||  
DB 1 VPSYSFIRAHDSVODLIA 20

RESULT 7  
US-09-290-049-19  
; Sequence 19, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: FDC98-01P2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; EARLIER FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 19  
LENGTH: 22  
TYPE: PRT  
ORGANISM: S. sobrinus  
US-09-290-049-19

Query Match 88.4%; Score 84; DB 16; Length 22;  
Best Local Similarity 88.9%; Pred. No. 2.3e-07;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHDSVODLI 18  
|||||  
DB 1 VPSYSFIRAHDSVODLI 18

RESULT 8  
US-09-290-049-18  
; Sequence 18, Application US/09290049

; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: FDC98-01P2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; EARLIER FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 18  
LENGTH: 22  
TYPE: PRT  
ORGANISM: S. downei  
US-09-290-049-18

Query Match 76.8%; Score 73; DB 16; Length 22;  
Best Local Similarity 78.9%; Pred. No. 2.1e-05;  
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHDSVODLIA 19  
||:|  
DB 1 VPNYVFIRAHDSVOTRIA 19

RESULT 9  
US-09-499-203-2  
; Sequence 2, Application US/09499203  
; GENERAL INFORMATION:  
; APPLICANT: KOSSMANN, Jens  
; APPLICANT: WELSH, Thomas  
; APPLICANT: GUANZ, Martin  
; APPLICANT: KNUTH, Karola  
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase  
; FILE REFERENCE: 147-196P  
; CURRENT APPLICATION NUMBER: US/09/499,203  
; CURRENT FILING DATE: 2000-02-08  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: Patentlin Ver. 2.1  
; SEQ ID NO 2  
LENGTH: 2057  
TYPE: PRT  
ORGANISM: Leuconostoc mesenteroides  
US-09-499-203-2

Query Match 69.5%; Score 66; DB 18; Length 2057;  
Best Local Similarity 61.1%; Pred. No. 0.099;  
Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VPSYSFIRAHDSVODLI 18  
:|:|:|:|:|:|:|:|  
DB 757 IPNYSFVRADHYDAQDPI 774

RESULT 10  
US-09-290-049-16  
; Sequence 16, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: FDC98-01P2A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; EARLIER FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13

EARLIER APPLICATION NUMBER: 60/115,142  
EARLIER FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO: 16  
LENGTH: 22  
TYPE: PRT  
ORGANISM: S. mutans  
US-09-290-049-16

Query Match 68.4%; Score 65; DB 16; Length 22;  
Best Local Similarity 76.5%; Pred. No. 0.00055;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 SYSFIRAHDESEVODLIA 19  
:|||||||||:11  
3 NYIFIRAHDESEVOTVIA 19

RESULT 11  
US-09-649-885-2  
Sequence 2, Application US/09649885  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Substitutes for Modified Starch and  
TITLE OF INVENTION: Latexes in Paper Manufacture  
FILE REFERENCE: 0358D2  
CURRENT APPLICATION NUMBER: US/09/649,885  
CURRENT FILING DATE: 2000-08-28  
PRIOR APPLICATION NUMBER: US 09/008,172  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/482,711  
PRIOR FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO: 2  
LENGTH: 1430  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-649-885-2

Query Match 68.4%; Score 65; DB 20; Length 1430;  
Best Local Similarity 76.5%; Pred. No. 0.095;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

3 SYSFIRAHDESEVODLIA 19  
:|||||||||:11  
Db 576 NYIFIRAHDESEVOTVIA 592

RESULT 12  
US-09-740-274-6  
Sequence 6, Application US/09740274  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Glucan-containing Compositions and Paper  
FILE REFERENCE: 0357CRD  
CURRENT APPLICATION NUMBER: US/09/740,274  
CURRENT FILING DATE: 2000-12-19  
PRIOR APPLICATION NUMBER: 09/210,361  
PRIOR FILING DATE: 1998-12-11  
PRIOR APPLICATION NUMBER: 09/007,999  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/478,704  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/009,620  
PRIOR FILING DATE: 1998-01-20  
PRIOR APPLICATION NUMBER: 08/485,243  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/008,172  
PRIOR FILING DATE: 1998-01-16

PRIOR APPLICATION NUMBER: 08/482,711  
PRIOR FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO: 6  
LENGTH: 1430  
TYPE: PRT  
ORGANISM: streptococcus mutans  
US-09-740-274-6

Query Match 68.4%; Score 65; DB 21; Length 1430;  
Best Local Similarity 76.5%; Pred. No. 0.095;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 SYSFIRAHDESEVODLIA 19  
:|||||||||:11  
Db 576 NYIFIRAHDESEVOTVIA 592

RESULT 13  
US-09-733-089-19824  
Sequence 19824, Application US/09733089  
GENERAL INFORMATION:  
APPLICANT: Dotson, Stanton B.  
APPLICANT: Kovalic, David K.  
APPLICANT: Liu, Jingdong  
APPLICANT: Lutfiyya, Linda L.  
APPLICANT: McIninch, James  
APPLICANT: Wu, Wei  
TITLE OF INVENTION: Nucleic Acid Molecules And other Molecules Associated With  
TITLE OF INVENTION: Transcription In Plants  
FILE REFERENCE: 38-21(15300)D  
CURRENT APPLICATION NUMBER: US/09/733,089  
CURRENT FILING DATE: 2000-12-11  
PRIOR APPLICATION NUMBER: US 09/474,435  
PRIOR FILING DATE: 1999-12-28  
PRIOR APPLICATION NUMBER: US 09/654,617  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: US 09/620,392  
PRIOR FILING DATE: 2000-07-19  
NUMBER OF SEQ ID NOS: 24143  
SEQ ID NO: 19824  
LENGTH: 263  
TYPE: PRT  
ORGANISM: Oryza sativa  
US-09-733-089-19824

Query Match 52.1%; Score 49.5; DB 21; Length 263;  
Best Local Similarity 45.0%; Pred. No. 6.7;  
Matches 9; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

QY 1 VPSYSFIRAHDS-EVODLIA 19  
:|||||:11  
Db 116 LPSFYPVRAHNGEISDTAA 135

RESULT 14  
US-09-816-660-19824  
Sequence 19824, Application US/09816660  
GENERAL INFORMATION:  
APPLICANT: Dotson, Stanton B.  
APPLICANT: Kovalic, David K.  
APPLICANT: Liu, Jingdong  
APPLICANT: Lutfiyya, Linda L.  
APPLICANT: McIninch, James  
APPLICANT: Wu, Wei  
TITLE OF INVENTION: Nucleic Acid Molecules And other Molecules Associated With  
TITLE OF INVENTION: Transcription In Plants  
FILE REFERENCE: 38-21(15300)D  
CURRENT APPLICATION NUMBER: US/09/816,660  
CURRENT FILING DATE: 2001-03-26

```

: PRIOR APPLICATION NUMBER: US 09/474,435
: PRIOR FILING DATE: 1999-12-28
: PRIOR APPLICATION NUMBER: US 09/654,617
: PRIOR FILING DATE: 2000-09-05
: PRIOR APPLICATION NUMBER: US 09/733,089
: PRIOR FILING DATE: 2000-12-11
: PRIOR APPLICATION NUMBER: ) US 09/684,016
: PRIOR FILING DATE: 2000-10-10
: PRIOR APPLICATION NUMBER: US 09/620,392
: PRIOR FILING DATE: 2000-07-19
: NUMBER OF SEQ ID NOS: 24143
: SEQ ID NO 19624
: LENGTH: 263
: TYPE: PRT
: ORGANISM: Oriza sativa
US-09-816-660-19824

```

```

/ Query Match 52.1%; Score 49.5; DB 22; Length 263;
Best Local Similarity 45.0%; Pred. No. 6.7;
Matches 9; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

1 VPSYSFIRAHDS-EVODLIA 19
:11: :1111: 1: 1
Db 116 LPSEFYVRAHDNGEISDTAA 135

```

```

RESULT 15
US-09-417-507-26458
: Sequence 26458, Application US/09417507
: GENERAL INFORMATION:
: APPLICANT: KEITH G. WEINSTOCK ET AL.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
: FILE REFERENCE: PATH99-10
: CURRENT APPLICATION NUMBER: US/09/417,507
: CURRENT FILING DATE: 1999-10-14
: NUMBER OF SEQ ID NOS: 44312
: SEQ ID NO 26458
: LENGTH: 541
: TYPE: PRT
: ORGANISM: A.fumigatus
US-09-417-507-26458

```

```

Query Match 49.5%; Score 47; DB 18; Length 541;
Best Local Similarity 47.6%; Pred. No. 46;
Matches 10; Conservative 4; Mismatches 5; Indels 2; Gaps 1;

1 VPSYSFIRAHDS-EVODLIA 19
:11: :11: 11: 11: 11:
204 VPOHSFLAHDQASVOSLLA 224

```

Search completed: March 27, 2002, 14:20:23  
 Job time: 1570 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:22:45 ; Search time 137.48 Seconds  
(without alignments)  
9.536 Million cell updates/sec

Title: US-09-290-049A-2

Perfect score: 95  
Sequence: 1 VPSYSTRAHDSVQDLIA 19

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 332938 seqs, 6899538 residues

Total number of hits satisfying chosen parameters: 332938

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PCR\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep1:\*  
7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep1:\*  
8: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	95	100.0	19	6	US-09-562-328-27
2	95	100.0	19	6	US-09-290-049A-2
3	95	100.0	23	6	US-09-562-328-44
4	91	95.8	22	6	US-09-562-328-28
5	91	95.8	22	6	US-09-290-049A-15
6	86	90.5	22	6	US-09-562-328-30
7	86	90.5	22	6	US-09-290-049A-17
8	84.5	88.9	20	6	US-09-562-328-26
9	84	88.4	22	6	US-09-562-328-35
10	84	88.4	22	6	US-09-290-049A-19
11	80	84.2	523	6	US-09-604-957-5
12	73	76.8	22	6	US-09-562-328-31
13	73	76.8	22	6	US-09-290-049A-18
14	67	70.5	584	6	US-09-604-957-6
15	66	69.5	535	6	US-09-604-957-7
16	66	69.5	1278	6	US-09-604-957-3
17	65	68.4	22	6	US-09-562-328-29
18	65	68.4	22	6	US-09-290-049A-16
19	65	68.4	545	6	US-09-604-957-4
20	47	49.5	274	6	US-09-675-784A-9123
21	44	46.3	327	8	US-60-356-051-2881
22	41	43.2	633	6	US-09-605-703B-440
23	40.5	42.6	389	6	US-09-708-427-10452
24	40.5	42.6	398	6	US-09-708-427-10451
25	40.5	42.6	414	6	US-09-708-427-10450

26	40.5	42.6	445	6	US-09-614-150-4236	Sequence 4236, Ap
27	40	42.1	547	6	US-09-943-857-2	Sequence 2, Appl
28	40	42.1	547	6	US-09-943-857-4	Sequence 4, Appl
29	40	42.1	547	6	US-09-943-857-6	Sequence 6, Appl
30	40	42.1	572	1	PCR-US02-03987-5626	Sequence 5626, Ap
31	40	42.1	572	6	US-09-815-242-5626	Sequence 5626, Ap
32	40	42.1	573	1	PCR-US02-03987-12247	Sequence 12247, A
33	40	42.1	573	1	PCR-US02-03987-12921	Sequence 12921, A
34	40	42.1	573	6	US-09-815-242-12247	Sequence 12247, A
35	40	42.1	573	6	US-09-815-242-12921	Sequence 12921, A
36	40	42.1	573	7	US-10-072-851-12247	Sequence 12247, A
37	40	42.1	573	7	US-10-072-851-12921	Sequence 12921, A
38	40	42.1	573	6	US-09-708-427-2857	Sequence 2857, Ap
39	40	42.1	737	6	US-09-708-427-2858	Sequence 2858, Ap
40	40	42.1	760	6	US-09-708-427-2858	Sequence 2858, Ap
41	40	42.1	783	6	US-09-708-427-2857	Sequence 2857, Ap
42	39	41.1	242	6	US-09-675-784A-9068	Sequence 9068, Ap
43	39	41.1	309	6	US-09-609-360C-37	Sequence 37, Appl
44	39	41.1	309	6	US-09-345-473E-37	Sequence 37, Appl
45	39	41.1	361	6	US-09-605-703B-1852	Sequence 1852, Ap

#### ALIGNMENTS

```
Result 1
US-09-562-328-27
; Sequence 27, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEES, ANDREW
; APPLICANT: SMITH, DANIEL J.
; APPLICANT: TAUBMAN, MARTIN A.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARRIES
; FILE REFERENCE: 04995.0046-01
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 09/288,965
; PRIOR FILING DATE: 1998-04-09
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Streptococcus sp.
US-09-562-328-27

Query Match      100.0%; Score 95; DB 6; Length 19;
Best Local Similarity 100.0%; Pred. No. 5.5e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 VPSYSTRAHDSVQDLIA 19
Db      1 VPSYSTRAHDSVQDLIA 19

RESULT 2
US-09-290-049A-2
; Sequence 2, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL CARRIES
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290,049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
```

SEQ ID NO 2  
LENGTH: 19  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: HDS peptide  
US-09-290-049a-2

Query Match 100.0%; Score 95; DB 6; Length 19;  
Best Local Similarity 100.0%; Pred. No. 5.5e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VPSSYFIRAHDSVEVDLIA 19  
|||||  
Db 1 VPSSYFIRAHDSVEVDLIA 19

RESULT 3  
US-09-562-328-44  
Sequence 44, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
APPLICANT: SMITH, DANIEL J.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 04995.0046-01  
CURRENT APPLICATION NUMBER: US/09/562,328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 44  
LENGTH: 23  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-44

Query Match 100.0%; Score 95; DB 6; Length 23;  
Best Local Similarity 100.0%; Pred. No. 7e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VPSSYFIRAHDSVEVDLIA 19  
|||||  
5 VPSSYFIRAHDSVEVDLIA 23

RESULT 4  
US-09-562-328-28  
Sequence 28, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
APPLICANT: SMITH, DANIEL J.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 04995.0046-01  
CURRENT APPLICATION NUMBER: US/09/562,328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 28  
LENGTH: 22  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-28

Query Match 95.8%; Score 91; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 3.2e-09;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 VPSSYFIRAHDSVEVDLI 18  
|||||  
Db 1 VPSSYFIRAHDSVEVDLI 18

RESULT 5  
US-09-290-049a-15  
Sequence 15, Application US/09290049A  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
FILE REFERENCE: 1564.1008-002  
CURRENT APPLICATION NUMBER: US/09/290,049A  
CURRENT FILING DATE: 1999-04-12  
PRIOR APPLICATION NUMBER: 60/081,550  
PRIOR FILING DATE: 1998-04-13  
PRIOR APPLICATION NUMBER: 60/115,142  
PRIOR FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15  
LENGTH: 22  
TYPE: PRT  
ORGANISM: S. mutans  
US-09-290-049a-15

Query Match 95.8%; Score 91; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 3.2e-09;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VPSSYFIRAHDSVEVDLI 18  
|||||  
Db 1 VPSSYFIRAHDSVEVDLI 18

RESULT 6  
US-09-562-328-30  
Sequence 30, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
APPLICANT: SMITH, DANIEL J.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 04995.0046-01  
CURRENT APPLICATION NUMBER: US/09/562,328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 30  
LENGTH: 22  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-30

Query Match 90.5%; Score 86; DB 6; Length 22;  
Best Local Similarity 94.4%; Pred. No. 2.4e-08;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VPSSYFIRAHDSVEVDLI 18  
|||||  
Db 1 VPSSYFIRAHDSVEVDLI 18

RESULT 7  
US-09-290-049a-17

```
; Sequence 17, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290.049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 22
; TYPE: PRT
; ORGANISM: S. downei
; US-09-290-049a-17
```

```
Query Match          90.5%; Score 86; DB 6; Length 22;
Best Local Similarity 94.4%; Pred. No. 2.4e-08;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 VPSYSFIRAHSEVODLI 18
    ||||| ||||| |||||
DB 1 VPSYSFIRAHSEVODLI 18
```

```
RESULT 8
US-09-562-328-26
; Sequence 26, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEES, ANDREW
; APPLICANT: TAUBMAN, MARTIN A.
; APPLICANT: SMITH, DANIEL J.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
; FILE REFERENCE: 04995.0046-01
; CURRENT APPLICATION NUMBER: US/09/562.328
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 09/288,965
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Streptococcus sp.
; US-562-328-26
```

```
Query Match          88.9%; Score 84.5; DB 6; Length 20;
Best Local Similarity 95.0%; Pred. No. 3.8e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
```

```
QY 1 VPSYSFIRAHSEVODLIA 19
    ||||| ||||| ||||| |||||
DB 1 VPSYSFIRAHSEVODLIA 20
```

```
RESULT 9
US-09-562-328-32
; Sequence 32, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEES, ANDREW
; APPLICANT: TAUBMAN, MARTIN A.
; APPLICANT: SMITH, DANIEL J.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
; FILE REFERENCE: 04995.0046-01
; CURRENT APPLICATION NUMBER: US/09/562.328
; CURRENT FILING DATE: 2000-05-01
```

```
; PRIOR APPLICATION NUMBER: 09/288,965
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Streptococcus sp.
; US-09-562-328-32
```

```
Query Match          88.4%; Score 84; DB 6; Length 22;
Best Local Similarity 88.9%; Pred. No. 5.2e-08;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 VPSYSFIRAHSEVODLI 18
    ||||| ||||| ||||| |||||
DB 1 VPSYSFIRAHSEVODII 18
```

```
RESULT 10
US-09-290-049a-19
; Sequence 19, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290.049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 22
; TYPE: PRT
; ORGANISM: S. sobrinus
; US-09-290-049a-19
```

```
Query Match          88.4%; Score 84; DB 6; Length 22;
Best Local Similarity 88.9%; Pred. No. 5.2e-08;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 VPSYSFIRAHSEVODLI 18
    ||||| ||||| ||||| |||||
DB 1 VPSYSFIRAHSEVODII 18
```

```
RESULT 11
US-09-604-957-5
; Sequence 5, Application US/09604957
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LOBBERT
; APPLICANT: RAHMOUTI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604.957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
```

— — — — —

Search completed: March 27, 2002, 14:22:46  
Job time: 1693 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:01:15 ; Search time 102.51 Seconds  
(without alignments)  
14.119 Million cell updates/sec

Title: US-09-290-049a-2

Perfect score: 95  
Sequence: 1 VPSYSTRAHDSVODLIA 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	95	100.0	1475	2	B3135
2	91	95.8	1375	2	JT0345
3	84	88.4	1592	2	A38175
4	80	84.2	1508	2	T31098
5	79	83.2	1290	2	JC5473
6	73	76.8	1365	2	A41483
7	65	68.4	1431	2	A45866
8	65	68.4	1449	2	T30857
9	65	68.4	1449	2	T30552
10	65	68.4	1518	2	A44811
11	64	67.4	1599	2	S22737
12	61	64.2	1577	2	T30858
13	46	48.4	575	2	S46329
14	44	46.3	327	2	D86474
15	43	45.3	51	2	G82455
16	43	45.3	214	1	RGBSCA
17	43	45.3	1260	2	A86323
18	42	44.2	431	2	S52583
19	42	44.2	563	2	P81888
20	42	44.2	563	2	DB1134
21	42	44.2	644	2	A64877
22	42	44.2	644	2	B85758
23	42	44.2	659	2	A64139
24	42	44.2	932	2	T40216
25	41.5	43.7	713	2	B64743
26	41.5	43.7	713	2	D85503
27	41	43.2	145	2	G69098
28	41	43.2	344	2	JH0511
29	41	43.2	597	2	F82935

30	41	43.2	1146	2	E70204	hypothetical prote
31	40.5	42.6	414	2	T06120	hypothetical prote
32	40	42.1	359	2	G82197	RstA1/RstA2 protei
33	40	42.1	431	2	B37802	crkA protein - Erw
34	40	42.1	534	2	S41735	cholesterol estera
35	40	42.1	548	2	S32615	triacylglycerol 11
36	40	42.1	549	2	JN0552	triacylglycerol 11
37	40	42.1	549	2	JN0551	triacylglycerol 11
38	40	42.1	572	2	B28474	phosphotransferase
39	40	42.1	783	2	T00782	probable anthranil
40	40	42.1	895	2	A86410	protein F3M18.22 f
41	40	42.1	1029	2	F86359	hypothetical prote
42	40	42.1	6359	2	T31679	bactiracin synthet
43	39.5	41.6	192	1	UCPEPP	cuticle protein pr
44	39	41.1	145	2	H84174	hypothetical prote
45	39	41.1	145	2	S07957	hypothetical prote

## ALIGNMENTS

RESULT 1  
B3135  
gtfB protein precursor - Streptococcus mutans  
C:Species: Streptococcus mutans  
C:Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 15-Oct-1999\*  
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
J. Bacteriol. 169, 4263-4270, 1987  
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
A:Reference number: A33135; MUID:87308013  
A:Accession: B33135  
A:Status: preliminary  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1475 <SHL>  
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
submitted to the Protein Sequence Database, September 1990  
A:Reference number: A33128  
A:Accession: A33128  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-171,173-641,'N',643-1475 <SH2>  
A:Experimental source: strain GS-5  
A:Superfamily: cpl repeat homology  
F:1096-1115/Domain: cpl repeat homology <CP1>  
F:1224-1243/Domain: cpl repeat homology <CP2>  
F:1289-1308/Domain: cpl repeat homology <CP3>  
F:1354-1373/Domain: cpl repeat homology <CP4>  
F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match 100.0%; Score 95; DB 2; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSYSTRAHDSVODLIA 19  
DB 552 VPSYSTRAHDSVODLIA 570  
RESULT 2  
JT0345  
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)  
N:Alternate names: sucrose 6-glucosyltransferase  
C:Species: Streptococcus mutans  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 24-Sep-1999  
A:Accession: JT0345; C33135  
R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.  
Gene 69, 101-109, 1988  
A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.  
A:Reference number: JT0345; MUID:89137980  
A:Accession: JT0345

A:Molecule type: DNA  
A:Residues: 1-1375 <UED>  
A:Experimental source: GS-5  
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
J. Bacteriol. 159, 4263-4270, 1997  
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
A:Reference number: A31135; MUID:87308013  
A:Accession: C31135  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-349 <SHI>  
A:Cross-references: GB:M17361  
C:Genetics:  
A:Gene: gtfC  
C:Function:  
A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans  
C:Superfamily: cpl repeat homology  
C:Keywords: duplication; glycosyltransferase; hexosyltransferase  
F:35-1375/Product: glycosyltransferase #status predicted <SIG>  
F:35-1375/Domain: signal sequence #status predicted <MAT>  
F:1126-1145/Domain: cpl repeat homology <CP1>  
F:1253-1272/Domain: cpl repeat homology <CP2>  
F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 95.8%; Score 91; DB 2; Length 1375;  
Best Local Similarity 100.0%; Pred. No. 4.7e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VPYSFIRAHDSFVODLI 18  
|||||  
Db 578 VPYSFIRAHDSFVODLI 595

RESULT 3  
A:8175  
glucosyltransferase precursor - Streptococcus sobrinus  
C:Species: Streptococcus sobrinus  
C>Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 15-Oct-1999  
C:Accession: A38175  
R:Abdo, H.; Msumura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.  
J. Bacteriol. 173, 989-996, 1991  
A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus  
A:Reference number: A38175; MUID:91123227  
A:Accession: A38175  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1592 <ABO>  
A:Cross-references: GB:D90213; NID:9217032; PIDN:BA14241.1; PID:d1014946; PID:g217033  
C:Superfamily: cpl repeat homology  
F:1093-1112/Domain: cpl repeat homology <CP1>  
F:1223-1241/Domain: cpl repeat homology <CP2>  
F:1287-1306/Domain: cpl repeat homology <CP3>  
F:1330-1351/Domain: cpl repeat homology <CP4>  
F:1352-1371/Domain: cpl repeat homology <CP5>  
F:1402-1420/Domain: cpl repeat homology <CP6>  
F:1465-1484/Domain: cpl repeat homology <CP7>  
F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 88.4%; Score 84; DB 2; Length 1592;  
Best Local Similarity 88.9%; Pred. No. 8.7e-06;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VPYSFIRAHDSFVODLI 18  
|||||  
Db 548 VPYSFIRAHDSFVODII 565

RESULT 4  
T31098  
Probable dextransucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides  
C:Species: Leuconostoc mesenteroides

C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
C:Accession: T31098  
R:Monchois, V.; Remaud-Simeon, M.; Monsan, P.; Willemot, R.M.  
FEMS Microbiol. Lett. 159, 307-315, 1998  
A:Title: Cloning and sequencing of a gene coding for an extracellular dextransucrase  
A:Reference number: 220981; MUID:98164374  
A:Accession: T31098  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1508 <MON>  
A:Cross-references: EMBL:AF030129; NID:92766611; PID:g2766612; PIDN:AAB95453.1  
A:Experimental source: strain NRRL B-1299  
C:Genetics:  
A:Gene: dsrB  
C:Function:  
A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 84.2%; Score 80; DB 2; Length 1508;  
Best Local Similarity 73.7%; Pred. No. 3.9e-05;  
Matches 14; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 VPYSFIRAHDSFVODLIA 19  
|||  
Db 634 IPNVSFVRAHDSFVQTYIA 652

RESULT 5  
JC5473  
dextransucrase (EC 2.4.1.5) - Leuconostoc mesenteroides  
C:Species: Leuconostoc mesenteroides  
C>Date: 07-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 29-Aug-1997  
C:Accession: JC5473  
R:Monchois, V.; Willemot, R.M.; Remaud-Simeon, M.; Croux, C.; Monsan, P.  
Gene 182, 23-32, 1996  
A:Title: Cloning and sequencing of a gene coding for a novel dextransucrase from Leuc  
A:Reference number: JC5473; MUID:97136686  
A:Accession: JC5473  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1290 <MON>  
A:Cross-references: GB:U38181  
C:Cross-references: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose  
A:Gene: dsrA  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:78-870/Domain: catalytic #status predicted <CAT>  
F:922-1290/Domain: glucan-binding #status predicted <GCB>

Query Match 83.2%; Score 79; DB 2; Length 1290;  
Best Local Similarity 83.3%; Pred. No. 4.9e-05;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 PYSFIRAHDSFVODLIA 19  
|  
Db 388 PYSFIRAHDSFVOTIIA 405

RESULT 6  
A41483  
glucosyltransferase (EC 2.4.1.-) gtfS precursor - Streptococcus sobrinus  
C:Species: Streptococcus sobrinus  
C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999  
C:Accession: A41483  
R:Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.  
Infect. Immun. 58, 2452-2458, 1990  
A:Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltr  
A:Reference number: A41483; MUID:90316665  
A:Accession: A41483  
A:Molecule type: DNA  
A:Residues: 1-1365 <GIL>

Query Match	68.48;	Score 65;	DB 2;	Length 1449;
Best Local Similarity	68.88;	Pred. NO.	0.014;	

RESULT 11  
S22737  
glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000  
C:Accession: S22737; S28810; B4481; S22727  
R:Jacques, N.

submitted to the EMBL Data Library, March 1992

A:Reference number: S22726

A:Accession: S22737

A:Molecule type: DNA

A:Residues: 1-1599 <JAC>

A:Cross-References: EMBL:211872; NID:947530; PIDN:CAA7898.1; PID:947531

A:Experimental source: ATCC 25975

R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.

J. Gen. Microbiol. 137, 2577-2593, 1991

A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes

A:Reference number: A44811; MUID:92148377

A:Accession: S28810

A:Molecule type: DNA

A:Residues: 1-51 <GIF>

A:Cross-References: EMBL:211873

C:Genetics:

A:Gene: gtfK

Superfamily: cpl repeat homology

Keywords: glycosyltransferase; hexosyltransferase

1456-1475/Domain: cpl repeat homology <CPR>

Query Match

Best Local Similarity 67.4%; Score 64; DB 2; Length 1599;

Matches 12; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 3 SYSFIRAHSEVODLIA 19

DB 574 TYLFVRAHSEVQTVIA 590

RESULT 12

glucosyltransferase - Streptococcus salivarius

C:Species: Streptococcus salivarius

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999

C:Accession: T30858

R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.

Infect. Immun. 63, 609-621, 1995

A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pr

A:Reference number: Z20909; MUID:95122197

A:Accession: T30858

A:Molecule type: DNA

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Residues: 1-1577 <SIM>

A:Cross-References: EMBL:L35928; NID:9662380; PID:9662381; PIDN:AA41413.1

C:Genetics:

A:Gene: gtfm

Query Match

Best Local Similarity 64.2%; Score 61; DB 2; Length 1577;

Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 3 SYSFIRAHSEVODLIA 19

DB 661 NYTFVRAHSEVQTVIA 677

RESULT 13

Intermediate filament protein - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 20-Jun-2000

C:Accession: S46329

R:Doedmont, H.; Riemer, D.; Ledger, N.; Weber, K.

EMBO J. 13, 2625-2638, 1994

A:Title: Eight genes and alternative RNA processing pathways generate an unexpectedly 1a

A:Reference number: S46326; MUID:94283388

A:Accession: S46329

A:Molecule type: DNA

A:Status: preliminary

A:Residues: 1-575 <DOD>

A:Cross-References: EMBL:X70833; NID:9312743; PIDN:CAA50181.1; PID:91848062

C:Genetics:

A:Introns: 31/1; 125/3; 220/3; 252/3; 290/3; 348/3; 445/1; 471/3; 563/3

C:Superfamily: Intermediate filament protein Av71

Query Match

Best Local Similarity 48.4%; Score 46; DB 2; Length 575;

Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 8 RAHDESEVODLIA 19

DB 248 RHDESEIHDLIA 259

RESULT 14

probable RING zinc finger protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: D86474

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Huizart, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Malt, R.; Marzla

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallio

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: D86474

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-327 <STO>

A:Cross-References: GB:AE005172; NID:911034939; PIDN:AMG27096.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match

Best Local Similarity 46.3%; Score 44; DB 2; Length 327;

Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 4 YSFIRAHSEVODL 17

DB 185 YPFIRHNDQMDL 198

RESULT 15

hypothetical protein VCA0471 [imported] - Vibrio cholerae (strain N16961 serogroup O1

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: G82455

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoti, I.; Sellers

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: AB2035; MUID:20406833

A:Accession: G82455

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-51 <HEI>

A:Cross-References: GB:AE004379; GB:AE003853; NID:99657865; PIDN:AAF96375.1; GSPDB:GN

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VCA0471

A:Map position: 2

Query Match 45.38; Score 43; DB 2; Length 51;  
 Best Local Similarity 58.88; Pred. No. 1.6;  
 Matches 10; Conservative 4; Mismatches 1; Indels 2; Gaps 1;  
 Oy 1 VPSYFIRAHDSDEVODL 17  
 ||:|||||:||:|  
 Db 15 VPATSAIR--NSEIRDL 29

Search completed: March 27, 2002, 14:01:16  
 Job time: 478 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen, Ltd.

OK protein - protein search, using sw model

Run on: March 27, 2002, 13:59:27 ; Search time 87.3 Seconds  
(without alignments)  
5.671 Million cell updates/sec

Title: US-09-290-049a-3

Perfect score: 112

Sequence: 1 TGAATTTCGOLIFRANGVYKG 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents-AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PTC05.COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	112	100.0	1475	3	US-09-007-999-2	Sequence 2, Appli
2	112	100.0	1475	4	US-09-210-361-2	Sequence 2, Appli
3	87	77.7	22	1	US-08-057-162B-3	Sequence 3, Appli
4	78	69.6	1375	4	US-09-210-361-4	Sequence 4, Appli
5	67	59.8	1430	4	US-09-008-172-2	Sequence 2, Appli
6	67	59.8	1430	4	US-09-210-361-6	Sequence 6, Appli
7	66	58.9	1577	2	US-08-793-824-2	Sequence 2, Appli
8	49.5	44.2	811	1	US-08-480-604A-7	Sequence 7, Appli
9	49.5	44.2	811	4	US-08-405-496A-7	Sequence 7, Appli
10	49.5	44.2	811	4	US-08-915-136-7	Sequence 7, Appli
11	49.5	44.2	812	1	US-08-480-604A-29	Sequence 29, Appli
12	49.5	44.2	812	4	US-08-915-136-29	Sequence 29, Appli
13	49.5	44.2	2710	1	US-08-480-604A-6	Sequence 6, Appli
14	49.5	44.2	2710	2	US-08-405-496A-6	Sequence 6, Appli
15	49.5	44.2	2710	2	US-08-915-136-6	Sequence 6, Appli
16	49.5	44.2	619	1	US-08-465-746-2	Sequence 2, Appli
17	46	41.1	619	1	US-08-214-164-2	Sequence 2, Appli
18	46	41.1	619	2	US-08-467-852A-3	Sequence 3, Appli
19	46	41.1	619	2	US-08-246-636-2	Sequence 2, Appli
20	46	41.1	619	2	US-08-247-491A-3	Sequence 3, Appli
21	46	41.1	619	2	US-08-319-795-2	Sequence 2, Appli
22	46	41.1	619	2	US-08-468-985-2	Sequence 2, Appli
23	46	41.1	619	3	US-08-312-949-2	Sequence 2, Appli
24	46	41.1	641	4	US-08-961-083-160	Sequence 160, App
25	46	41.1	648	1	US-08-072-070-2	Sequence 2, Appli
26	46	41.1	648	1	US-08-469-434-2	Sequence 2, Appli
27	46	41.1	648	1	US-08-214-222-2	Sequence 2, Appli

28	46	41.1	648	2	US-08-467-852A-2	Sequence 2, Appli
29	46	41.1	648	2	US-08-468-718-2	Sequence 2, Appli
30	46	41.1	648	2	US-08-247-491A-2	Sequence 2, Appli
31	46	41.1	648	3	US-08-446-201-3	Sequence 3, Appli
32	46	41.1	695	1	US-08-127-499A-23	Sequence 23, Appli
33	46	41.1	695	1	US-08-482-847-23	Sequence 23, Appli
34	43	38.4	1477	1	US-08-038-682-4	Sequence 4, Appli
35	43	38.4	1477	1	US-08-302-832-4	Sequence 4, Appli
36	43	38.4	1477	2	US-08-530-198-4	Sequence 4, Appli
37	43	38.4	1477	2	US-08-469-880-4	Sequence 4, Appli
38	43	38.4	1477	2	US-08-728-470-4	Sequence 4, Appli
39	43	38.4	1477	2	US-08-617-697-4	Sequence 4, Appli
40	43	38.4	1477	2	US-08-719-641-4	Sequence 4, Appli
41	40	35.7	265	1	US-08-461-731-2	Sequence 2, Appli
42	40	35.7	265	1	US-09-002-072B-2	Sequence 2, Appli
43	40	35.7	265	4	US-09-314-199-2	Sequence 2, Appli
44	40	35.7	265	5	PCR-US94-10465-3	Sequence 3, Appli
45	40	35.7	654	1	US-08-392-828C-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
; US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0356D
; CURRENT APPLICATION NUMBER: US/09/007, 999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478, 704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; US-09-007-999-2

Query Match 100.0%, Score 112, DB 3, Length 1475;
Best Local Similarity 100.0%, Pred. No. 9, 3e-11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAATTTCGOLIFRANGVYKG 22
DB 1300 TGAATTTCGOLIFRANGVYKG 1321

RESULT 2
; US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210, 361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007, 999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478, 704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 08/009, 620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485, 243
; EARLIER FILING DATE: 1995-06-07
```

EARLIER APPLICATION NUMBER: 09/008,172  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/482,711  
EARLIER FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1475  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-210-361-2

Query Match 100.0%; Score 112; DB 4; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 9,3e-11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TGARTINGQLLYFRANGVQYKG 22  
|||||  
1300 TGARTINGQLLYFRANGVQYKG 1321

RESULT 3  
US-08-057-162B-3  
Sequence 3, Application US/08057162B  
Patent No. 5686075  
GENERAL INFORMATION:  
APPLICANT: Taubman, Martin A.  
APPLICANT: Smith, Daniel J.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINE FOR DENTAL CARRIES  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/057,162B  
FILING DATE: 30-APR-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/877,295  
FILING DATE: 01-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wagner, Richard W.  
REGISTRATION NUMBER: 34,480  
REFERENCE/DOCKET NUMBER: FDC92-01A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-057-162B-3

Query Match 77.7%; Score 87; DB 1; Length 22;  
Best Local Similarity 77.3%; Pred. No. 1.1e-08;  
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
1 TGARTINGQLLYFRANGVQYKG 22  
|||||

Db 1 TGAQTIKGQKLYFKANGQYQYKG 22

RESULT 4  
US-09-210-361-4  
Sequence 4, Application US/09210361  
Patent No. 6284479  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Substitutes for Modified Starches and  
FILE REFERENCE: 0357CR  
CURRENT APPLICATION NUMBER: US/09/210,361  
CURRENT FILING DATE: 1998-12-11  
EARLIER APPLICATION NUMBER: 09/007,999  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/478,704  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 09/009,620  
EARLIER FILING DATE: 1998-01-20  
EARLIER APPLICATION NUMBER: 08/485,243  
EARLIER FILING DATE: 1995-06-07  
EARLIER APPLICATION NUMBER: 09/008,172  
EARLIER FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/482,711  
EARLIER FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 1375  
TYPE: PRT  
ORGANISM: streptococcus mutans  
US-09-210-361-4

Query Match 69.6%; Score 78; DB 4; Length 1375;  
Best Local Similarity 68.2%; Pred. No. 6.1e-05;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGARTINGQLLYFRANGVQYKG 22  
|||

Db 1264 TGTVTENGRLYFKPANGVQAKG 1285  
RESULT 5  
US-09-008-172-2  
Sequence 2, Application US/09008172  
Patent No. 6127602  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Substitutes for Modified Starch and  
FILE REFERENCE: 0358D  
CURRENT APPLICATION NUMBER: US/09/008,172  
CURRENT FILING DATE: 1998-01-16  
EARLIER APPLICATION NUMBER: 08/482,711  
EARLIER FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1430  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-008-172-2

Query Match 59.8%; Score 67; DB 3; Length 1430;  
Best Local Similarity 59.1%; Pred. No. 0.005;  
Matches 13; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
QY 1 TGARTINGQLLYFRANGVQYKG 22  
|||

Db 1332 TGSOTIAGKKLYFASDGKQYQYKG 1353



```

RESULT 6
US-09-210-361-6
: Sequence 6, Application US/09210361
: Patent No. 6284479
: GENERAL INFORMATION:
: APPLICANT: Nichols, Scott E.
: TITLE OF INVENTION: Substitutes for Modified Starches and
: TITLE OF INVENTION: Latexes in Paper Manufacture
: FILE REFERENCE: 0357CR
: CURRENT APPLICATION NUMBER: US/09/210.361
: CURRENT FILING DATE: 1998-12-11
: EARLIER APPLICATION NUMBER: 09/007,999
: EARLIER FILING DATE: 1998-01-16
: EARLIER APPLICATION NUMBER: 08/478,704
: EARLIER FILING DATE: 1995-06-07
: EARLIER APPLICATION NUMBER: 09/009,620
: EARLIER FILING DATE: 1998-01-20
: EARLIER APPLICATION NUMBER: 08/485,243
: EARLIER FILING DATE: 1995-06-07
: EARLIER APPLICATION NUMBER: 09/008,172
: EARLIER FILING DATE: 1998-01-16
: EARLIER APPLICATION NUMBER: 08/482,711
: EARLIER FILING DATE: 1995-06-07
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 6
: LENGTH: 1430
: TYPE: PRF
: ORGANISM: streptococcus mutans
US-09-210-361-6

```

OY 1 TGAARTINGOLLYFRANGYOVKG 22  
11::11 1: 111 :1 1111  
Db 1332 TGSOFTIAGKRLFYASDGROVKG 1353

Matches 13; Conservative 5; Mismatches 4; Indels 0; Gaps 0

RESULT 7  
US-08-793-824-2  
Sequence 2, Application US/08793824  
Patent No. 5981838  
GENERAL INFORMATION:  
APPLICANT: Simpson, Christine Lynn  
APPLICANT: Giffard, Phillip Morrison  
APPLICANT: Jacques, Nicholas Anthony  
TITLE OF INVENTION: Genetic Manipulation of Plants to  
Title of Invention: Increase stored Carbohydrates  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Griffith Hack & Co  
STREET: Level 8, 168 Walker Street  
CITY: No. 5981838th Sydney  
STATE: New South Wales  
COUNTRY: Australia  
ZIP: 2060  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/793.824  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PM7643  
FILING DATE: 24-AUG-1994

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 61 2 9957 5944  
TELEFAX: 61 2 957 6288  
TELEX: 26547  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1577 amino acids  
type: amino acid  
STRANDEDNESS:  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus salivarius  
US-08-793-824-2

Query Match	58.9%	Score 66	DB 2	Length 1577
Best Local Similarity	54.5%	Pred. No. 0.0083		
Matches 12	Conservative	4	Mismatches 6	Indels 0
			Gaps	0

```
QY 1 TGARTINGQLLYFRANGVQVKG 22
    || : || : || : || ||||
Db 1469 TGLQININKVYFEGSNGAQVKG 1490
```

RESULT 8  
US-08-480-604A-7  
Sequence 7, Application US/08480604A  
Patent No. 5736139  
GENERAL INFORMATION:  
APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BRUCE S.  
APPLICANT: PADHEE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MONTGOMERY & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA  
Zip: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,604A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/422,711  
FILING DATE: 14-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,496  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:

```

; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 811 amino acids
; TYPE: amino acid

```

STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-915-136-7

Query Match 44.2%; Score 49.5; DB 4; Length 811;  
Best Local Similarity 52.2%; Pred. No. 2.6;  
Matches 12: Conservative 2; Mismatches 8; Indels 1; Gaps 1;

Oy 1 TGARTINGOLLYFRAN-GVQVKG 22  
Db 587 TGLRTIDGKKYFNTNTAVATG 609

RESULT 11  
US-08-480-604A-29  
Sequence 29, Application US/08480604A  
Patent No. 5736139  
GENERAL INFORMATION:  
APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,604A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/422,711  
FILING DATE: 14-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,496  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01763  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 812 amino acids  
TYPE: amino acid

STRANDEDNESS: unknown  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-480-604A-29

Query Match 44.2%; Score 49.5; DB 1; Length 812;  
Best Local Similarity 52.2%; Pred. No. 2.6;  
Matches 12: Conservative 2; Mismatches 8; Indels 1; Gaps 1;

Oy 1 TGARTINGOLLYFRAN-GVQVKG 22  
Db 588 TGLRTIDGKKYFNTNTAVATG 610

RESULT 12  
US-08-915-136-29  
Sequence 29, Application US/08915136  
Patent No. 6290960  
GENERAL INFORMATION:  
APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/915,136  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/480,604  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,496  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01763  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 812 amino acids  
TYPE: amino acid

```

; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-915-136-29

```

Query Match	44.28;	Score 49.5;	DB 4;	Length 812;
Best Local Similarity	52.28;	Pred. No. 2.6;		
Matches 12; Conservative	2;	Mismatches 8;	Indels 1;	Gaps 1

Qy 1 TGARTINGQLLYFRAN-GVQVKG 22  
|||: |||  
Db 588 TGLRTIDGKKYYFNTNTAAVATG 610

RESULT 13  
US-08-480-604A-6  
Sequence 6, Application US/08480604A  
Patent No. 5736139  
GENERAL INFORMATION:  
INVENTOR: TOWN

APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE  
TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:

```

COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:

```

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/422,711  
FILING DATE: 14-APR-1995

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIORITY APPLICATION DATA:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:

ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01763

INFORMATION FOR SEQ ID NO: 6  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2710 amino acids  
 TYPE: amino acid

```

;      TOPOLOGY:  linear
;      MOLECULE TYPE:  protein
US-08-480-604A-6

```

Query Match	44.28;	Score 49.5;	DB 1;	Length 2710;
Best Local Similarity	52.28;	Pred. No. 11;		
Matches 12;	Conservative 2;	Mismatches 8;	Indels 1;	Gaps 1;

```
QY      1 TGARTINGQLLYERAN-GVQVKG 22
        |||||: || | |||
Db      2460 TGLRTIDGKKYYFNTNTAVAVTG 2482
```

RESULT 14  
US-08-405-496A-6  
; Sequence 6, Application US/08405496A  
; Patent No. 5919665  
; GENERAL INFORMATION:  
ADDENDUM: CITATIONS TAKEN 1

APPLICANT: WILLIAMS, JAMES A.  
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
TITLE OF INVENTION: NEUROTOXIN  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: USA

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/4405.496A  
FILING DATE: 16-MAR-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329.154

FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321

FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INCOLIA, DIANE E.

TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:

```

;      TOPOLOGY:  linear
;      MOLECULE TYPE:  protein
US-08-405-496A-6

```

Query Match	44.2%	Score 49.5;	DB 2;	length 2710;
Best Local Similarity	52.2%;	Pred. NO. 11;		
Matches	12;	Conservative	2;	Mismatches 8;
				Indels 1;
				Gaps 1.

Db 2460 TGLRTIDGKKYYFNTNTAVATG 2482

## RESULT 15

US-08-915-136-6

Sequence 6, Application US/08915136  
Patent No. 6290960

## GENERAL INFORMATION:

APPLICANT: KINK, JOHN A.

APPLICANT: THALLEY, BRUCE S.

APPLICANT: PADHIE, NISHA V.

APPLICANT: FIRCA, JOSEPH R.

APPLICANT: STAFORD, DOUGLAS C.

TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESSES:

ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: UNITED STATES OF AMERICA

ZIP: 94104

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/915,136

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/480,604

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/405,496

FILING DATE: 16-MAR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/329,154

FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/161,907

FILING DATE: 02-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989

PRIOR APPLICATION DATA:

ATTORNEY/AGENT INFORMATION:

NAME: INGOLIA, DIANE E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPHD-01763

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 2710 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-915-136-6

Query Match 44.2%; Score 49.5; DB 4; Length 2710;

Best Local Similarity 52.2%; Pred. No. 11;

Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

OY 1 TGAATNGQLYFRAN-GVQVKG 22

Db 2460 TGLRTIDGKKYYFNTNTAVATG 2482

Search completed: March 27, 2002, 13:59:28  
Job time: 561 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:20:23 ; Search time 1139.61 seconds  
(without alignments)  
5.360 Million cell updates/sec

Title: US-09-290-049a-3  
Perfect score: 112  
Sequence: 1 TCGARTINGOLLYFRANGVQVKG 22

Scoring table: BLOSOM62  
Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

Total number of hits satisfying chosen parameters: 3148936

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main: \*  
1: /cgn2\_6/ptodata/2/paa/PCRTUS\_COMB.pep: \*  
2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep: \*  
3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep: \*  
4: /cgn2\_6/ptodata/2/paa/US08\_COMB.pep: \*  
5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep: \*  
6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep: \*  
7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep: \*  
8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep: \*  
9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep: \*  
10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep: \*  
11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep: \*  
12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep: \*  
13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep: \*  
14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep: \*  
15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep: \*  
16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep: \*  
17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep: \*  
18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep: \*  
19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep: \*  
20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep: \*  
21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep: \*  
22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep: \*  
23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep: \*  
24: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	112	100.0	22	16	US-09-290-049-3
2	112	100.0	1475	19	US-09-557-848-2
3	112	100.0	1475	21	US-09-740-274-2
4	87	77.7	22	1	PCT-US93-04094-3
5	87	77.7	22	3	US-07-877-295-2
6	87	77.7	22	4	US-08-057-162-3
7	87	77.7	22	4	US-08-057-162A-3
8	87	77.7	22	16	US-09-288-965-3
9	78	69.6	1375	21	US-09-740-274-4

	10	67	59.8	1430	20	US-09-649-885-2	Sequence 2, Appli
	11	67	59.8 <td>1430</td> <td>21</td> <td>US-09-740-274-6<td>Sequence 6, Appli</td></td>	1430	21	US-09-740-274-6 <td>Sequence 6, Appli</td>	Sequence 6, Appli
	12	65	58.0 <td>2057</td> <td>18</td> <td>US-09-499-203-2<td>Sequence 2, Appli</td></td>	2057	18	US-09-499-203-2 <td>Sequence 2, Appli</td>	Sequence 2, Appli
	13	49.5	44.2 <td>320</td> <td>18</td> <td>US-09-446-269-1<td>Sequence 1, Appli</td></td>	320	18	US-09-446-269-1 <td>Sequence 1, Appli</td>	Sequence 1, Appli
	14	49.5	44.2 <td>457</td> <td>18</td> <td>US-09-446-269-3<td>Sequence 3, Appli</td></td>	457	18	US-09-446-269-3 <td>Sequence 3, Appli</td>	Sequence 3, Appli
	15	49.5	44.2 <td>811</td> <td>1</td> <td>PCT-US97-15394-7<td>Sequence 7, Appli</td></td>	811	1	PCT-US97-15394-7 <td>Sequence 7, Appli</td>	Sequence 7, Appli
	16	49.5	44.2 <td>811</td> <td>8</td> <td>US-08-405-496-7<td>Sequence 7, Appli</td></td>	811	8	US-08-405-496-7 <td>Sequence 7, Appli</td>	Sequence 7, Appli
	17	49.5	44.2 <td>811</td> <td>8</td> <td>US-08-422-711-7<td>Sequence 7, Appli</td></td>	811	8	US-08-422-711-7 <td>Sequence 7, Appli</td>	Sequence 7, Appli
	18	49.5	44.2 <td>811</td> <td>8</td> <td>US-08-480-604-7<td>Sequence 7, Appli</td></td>	811	8	US-08-480-604-7 <td>Sequence 7, Appli</td>	Sequence 7, Appli
	19	49.5	44.2 <td>811</td> <td>13</td> <td>US-08-704-159-7<td>Sequence 7, Appli</td></td>	811	13	US-08-704-159-7 <td>Sequence 7, Appli</td>	Sequence 7, Appli
	20	49.5	44.2 <td>811</td> <td>13</td> <td>US-08-957-310-7<td>Sequence 7, Appli</td></td>	811	13	US-08-957-310-7 <td>Sequence 7, Appli</td>	Sequence 7, Appli
	21	49.5	44.2 <td>811</td> <td>14</td> <td>US-09-084-517-7<td>Sequence 7, Appli</td></td>	811	14	US-09-084-517-7 <td>Sequence 7, Appli</td>	Sequence 7, Appli
	22	49.5	44.2 <td>811</td> <td>19</td> <td>US-09-587-198-7<td>Sequence 7, Appli</td></td>	811	19	US-09-587-198-7 <td>Sequence 7, Appli</td>	Sequence 7, Appli
	23	49.5	44.2 <td>812</td> <td>8</td> <td>US-08-422-711-29<td>Sequence 29, Appli</td></td>	812	8	US-08-422-711-29 <td>Sequence 29, Appli</td>	Sequence 29, Appli
	24	49.5	44.2 <td>812</td> <td>8</td> <td>US-08-480-604-29<td>Sequence 29, Appli</td></td>	812	8	US-08-480-604-29 <td>Sequence 29, Appli</td>	Sequence 29, Appli
	25	49.5	44.2 <td>812</td> <td>14</td> <td>US-09-084-517-29<td>Sequence 29, Appli</td></td>	812	14	US-09-084-517-29 <td>Sequence 29, Appli</td>	Sequence 29, Appli
	26	49.5	44.2 <td>862</td> <td>18</td> <td>US-09-446-269-4<td>Sequence 4, Appli</td></td>	862	18	US-09-446-269-4 <td>Sequence 4, Appli</td>	Sequence 4, Appli
	27	49.5	44.2 <td>866</td> <td>19</td> <td>US-09-545-772-2<td>Sequence 2, Appli</td></td>	866	19	US-09-545-772-2 <td>Sequence 2, Appli</td>	Sequence 2, Appli
	28	49.5	44.2 <td>866</td> <td>19</td> <td>US-09-545-773-2<td>Sequence 2, Appli</td></td>	866	19	US-09-545-773-2 <td>Sequence 2, Appli</td>	Sequence 2, Appli
	29	49.5	44.2 <td>2710</td> <td>1</td> <td>PCT-US97-15394-6<td>Sequence 6, Appli</td></td>	2710	1	PCT-US97-15394-6 <td>Sequence 6, Appli</td>	Sequence 6, Appli
	30	49.5	44.2 <td>2710</td> <td>8</td> <td>US-08-405-496-6<td>Sequence 6, Appli</td></td>	2710	8	US-08-405-496-6 <td>Sequence 6, Appli</td>	Sequence 6, Appli
	31	49.5	44.2 <td>2710</td> <td>8</td> <td>US-08-422-711-6<td>Sequence 6, Appli</td></td>	2710	8	US-08-422-711-6 <td>Sequence 6, Appli</td>	Sequence 6, Appli
	32	49.5	44.2 <td>2710</td> <td>8</td> <td>US-08-480-604-6<td>Sequence 6, Appli</td></td>	2710	8	US-08-480-604-6 <td>Sequence 6, Appli</td>	Sequence 6, Appli
	33	49.5	44.2 <td>2710</td> <td>11</td> <td>US-08-704-159-6<td>Sequence 6, Appli</td></td>	2710	11	US-08-704-159-6 <td>Sequence 6, Appli</td>	Sequence 6, Appli
	34	49.5	44.2 <td>2710</td> <td>13</td> <td>US-08-957-310-6<td>Sequence 6, Appli</td></td>	2710	13	US-08-957-310-6 <td>Sequence 6, Appli</td>	Sequence 6, Appli
	35	49.5	44.2 <td>2710</td> <td>14</td> <td>US-09-084-517-6<td>Sequence 6, Appli</td></td>	2710	14	US-09-084-517-6 <td>Sequence 6, Appli</td>	Sequence 6, Appli
	36	49.5	44.2 <td>2710</td> <td>19</td> <td>US-09-587-198-6<td>Sequence 6, Appli</td></td>	2710	19	US-09-587-198-6 <td>Sequence 6, Appli</td>	Sequence 6, Appli
	37	47	42.0 <td>785</td> <td>18</td> <td>US-09-488-725A-2538<td>Sequence 2538, Ap</td></td>	785	18	US-09-488-725A-2538 <td>Sequence 2538, Ap</td>	Sequence 2538, Ap
	38	47	42.0 <td>785</td> <td>19</td> <td>US-09-538-092-872<td>Sequence 872, App</td></td>	785	19	US-09-538-092-872 <td>Sequence 872, App</td>	Sequence 872, App
	39	47	42.0 <td>794</td> <td>18</td> <td>US-09-488-725A-6110<td>Sequence 6110, Ap</td></td>	794	18	US-09-488-725A-6110 <td>Sequence 6110, Ap</td>	Sequence 6110, Ap
	40	46	41.1 <td>619</td> <td>19</td> <td>US-09-502-527-4<td>Sequence 4, Appli</td></td>	619	19	US-09-502-527-4 <td>Sequence 4, Appli</td>	Sequence 4, Appli
	41	46	41.1 <td>619</td> <td>22</td> <td>US-09-882-774-1<td>Sequence 1, Appli</td></td>	619	22	US-09-882-774-1 <td>Sequence 1, Appli</td>	Sequence 1, Appli
	42	46	41.1 <td>641</td> <td>19</td> <td>US-09-536-784-160<td>Sequence 160, App</td></td>	641	19	US-09-536-784-160 <td>Sequence 160, App</td>	Sequence 160, App
	43	46	41.1 <td>641</td> <td>21</td> <td>US-09-765-271-160<td>Sequence 160, App</td></td>	641	21	US-09-765-271-160 <td>Sequence 160, App</td>	Sequence 160, App
	44	46	41.1 <td>641</td> <td>21</td> <td>US-09-765-272-160<td>Sequence 160, App</td></td>	641	21	US-09-765-272-160 <td>Sequence 160, App</td>	Sequence 160, App
	45	46	41.1 <td>648</td> <td>6</td> <td>US-08-247-491-2<td>Sequence 2, Appli</td></td>	648	6	US-08-247-491-2 <td>Sequence 2, Appli</td>	Sequence 2, Appli

#### ALIGNMENTS

RESULT 1  
US-09-290-049-3  
Sequence 3, Application US-09290049  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
FILE REFERENCE: FDC98-0162A  
CURRENT APPLICATION NUMBER: US-09/290,049  
CURRENT FILING DATE: 1999-04-12  
EARLIER APPLICATION NUMBER: 60/081,550  
EARLIER FILING DATE: 1998-04-13  
EARLIER APPLICATION NUMBER: 60/115,142  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 3  
LENGTH: 22  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: GJB peptide  
US-09-290-049-3

Query Match 100.0%; Score 112; DB 16; Length 22;  
Best Local Similarity 100.0%; Pred. No. 1.7e-11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TCGARTINGOLLYFRANGVQVKG 22  
|||||

Db 1 TGARTINGQLYFRANGVQVG 22

## RESULT 2

US-09-557-848-2  
; Sequence 2, Application US/09557848  
; GENERAL INFORMATION:  
; APPLICANT: NICHOLS, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; TITLE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0356D2  
; CURRENT APPLICATION NUMBER: US/09/557,848  
; CURRENT FILING DATE: 2000-04-26  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-557-848-2

Query Match 100.0%; Score 112; DB 19; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 2,1e-09;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGARTINGQLYFRANGVQVG 22  
|||||  
DB 1300 TGARTINGQLYFRANGVQVG 1321

## RESULT 3

US-09-740-274-2  
; Sequence 2, Application US/09740274  
; GENERAL INFORMATION:  
; APPLICANT: NICHOLS, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-740-274-2

Query Match 100.0%; Score 112; DB 21; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 2,1e-09;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGARTINGQLYFRANGVQVG 22  
|||||

Db 1300 TGARTINGQLYFRANGVQVG 1321

## RESULT 4

PCT-US93-04094-3  
; Sequence 3, Application PC/TUS9304094  
; GENERAL INFORMATION:  
; APPLICANT: Forsyth Dental Infirmary for Children  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR  
; TITLE OF INVENTION: DENTAL  
; TITLE OF INVENTION: CARIES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Millitia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/04094  
; FILING DATE: 19930430  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia.  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: FDC92-01A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 861-6240  
; TELEFAX: (617) 861-9540  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 22 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
PCT-US93-04094-3

Query Match 77.7%; Score 87; DB 1; Length 22;  
Best Local Similarity 77.3%; Pred. No. 3,1e-07;  
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGARTINGQLYFRANGVQVG 22  
|||||  
DB 1 TGAQTIKGOKLYFRANGVQVG 22

## RESULT 5

US-07-877-295-2  
; Sequence 2, Application US/07877295  
; GENERAL INFORMATION:  
; APPLICANT: Taudman, Martin A.  
; APPLICANT: Smith, Daniel J.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARIES  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Millitia Drive  
; CITY: Lexington  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible



OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/877,295  
FILING DATE: 19920501  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Wagner, Richard W.  
REGISTRATION NUMBER: 34,480  
REFERENCE/DOCKET NUMBER: FDC92-01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-877-295-2

Query Match 77.7%; Score 87; DB 3; Length 22;  
Best Local Similarity 77.3%; Pred. No. 3,1e-07;  
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGAQTINGQLYFRANGVOYVG 22  
DB 1 TGAQTINGKOKLYFRANGGOYVG 22

RESULT 6  
US-08-057-162-3  
Sequence 3, Application US/08057162  
GENERAL INFORMATION:  
APPLICANT: Taudman, Martin A.  
APPLICANT: Smith, Daniel J.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR  
TITLE OF INVENTION: DENTAL  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/057,162  
FILING DATE: 19930430  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/877,295  
FILING DATE: 01-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wagner, Richard W.  
REGISTRATION NUMBER: 34,480  
REFERENCE/DOCKET NUMBER: FDC92-01A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-057-162-3

Query Match 77.7%; Score 87; DB 4; Length 22;  
Best Local Similarity 77.3%; Pred. No. 3,1e-07;  
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGAQTINGQLYFRANGVOYVG 22  
DB 1 TGAQTINGKOKLYFRANGGOYVG 22

RESULT 7  
US-08-057-162A-3  
Sequence 3, Application US/08057162A  
GENERAL INFORMATION:  
APPLICANT: Taudman, Martin A.  
APPLICANT: Smith, Daniel J.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINE FOR DENTAL  
TITLE OF INVENTION: CARIES  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/057,162A  
FILING DATE: 30-APR-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/877,295  
FILING DATE: 01-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wagner, Richard W.  
REGISTRATION NUMBER: 34,480  
REFERENCE/DOCKET NUMBER: FDC92-01A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-057-162A-3

Query Match 77.7%; Score 87; DB 4; Length 22;  
Best Local Similarity 77.3%; Pred. No. 3,1e-07;  
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGAQTINGQLYFRANGVOYVG 22  
DB 1 TGAQTINGKOKLYFRANGGOYVG 22

RESULT 8  
US-09-288-965-3  
Sequence 3, Application US/09288965  
GENERAL INFORMATION:  
APPLICANT: Lees, Andrew

APPLICANT: Taubman, Martin A.  
APPLICANT: Smith, Daniel J.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 04995.0046-00000  
CURRENT APPLICATION NUMBER: US/09/288,965  
CURRENT FILING DATE: 1999-04-09  
EARLIER APPLICATION NUMBER: 60/081,315  
EARLIER FILING DATE: 1998-04-10  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3  
LENGTH: 22  
TYPE: PRT  
ORGANISM: Streptococcus  
FEATURE:  
OTHER INFORMATION: may be S. mutans, S. rattus, S. cricetus or S. sobinus  
US-09-288-965-3

Query Match 77.7%: Score 87; DB 16; Length 22;  
Best Local Similarity 77.3%: Pred. No. 3.1e-07;  
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGARTINGQLYFRANGVQVG 22  
||:|||||:|||||:|||||  
DB 1 TGAQTIKGQKLYFRANGQOVKG 22

RESULT 9  
US-09-740-274-4  
Sequence 4, Application US/09740274  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Glucan-containing Compositions and Paper  
FILE REFERENCE: 0357CRD  
CURRENT APPLICATION NUMBER: US/09/740,274  
CURRENT FILING DATE: 2000-12-19  
PRIOR APPLICATION NUMBER: 09/210,361  
PRIOR FILING DATE: 1998-12-11  
PRIOR APPLICATION NUMBER: 09/007,999  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/478,704  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/009,620  
PRIOR FILING DATE: 1998-01-20  
PRIOR APPLICATION NUMBER: 08/485,243  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/008,172  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/482,711  
PRIOR FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 1375  
TYPE: PRT  
ORGANISM: streptococcus mutans  
US-09-740-274-4

Query Match 69.6%: Score 78; DB 21; Length 1375;  
Best Local Similarity 68.2%: Pred. No. 0.0012;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGARTINGQLYFRANGVQVG 22  
||:|||||:|||||:|||||  
DB 1264 TGTVTFGQRLYFKPBGVQAKG 1285

RESULT 10  
US-09-649-885-2  
Sequence 2, Application US/09649885  
GENERAL INFORMATION:

APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Substitutes for Modified Starch and  
TITLE OF INVENTION: Latexes in Paper Manufacture  
FILE REFERENCE: 0358B2  
CURRENT APPLICATION NUMBER: US/09/649,885.  
CURRENT FILING DATE: 2000-08-28  
PRIOR APPLICATION NUMBER: US 09/008,172  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/482,711  
PRIOR FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1430  
TYPE: PRT  
ORGANISM: Streptococcus mutans  
US-09-649-885-2

Query Match 59.8%: Score 67; DB 20; Length 1430;  
Best Local Similarity 59.1%: Pred. No. 0.091;  
Matches 13; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGARTINGQLYFRANGVQVG 22  
||:|||||:|||||:|||||  
DB 1332 TGSQTIAKKLYFASDGKQVKG 1353

RESULT 11  
US-09-740-274-6  
Sequence 6, Application US/09740274  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Glucan-containing Compositions and Paper  
FILE REFERENCE: 0357CRD  
CURRENT APPLICATION NUMBER: US/09/740,274  
CURRENT FILING DATE: 2000-12-19  
PRIOR APPLICATION NUMBER: 09/210,361  
PRIOR FILING DATE: 1998-12-11  
PRIOR APPLICATION NUMBER: 09/007,999  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/478,704  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/009,620  
PRIOR FILING DATE: 1998-01-20  
PRIOR APPLICATION NUMBER: 08/485,243  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 09/008,172  
PRIOR FILING DATE: 1998-01-16  
PRIOR APPLICATION NUMBER: 08/482,711  
PRIOR FILING DATE: 1995-06-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 6  
LENGTH: 1430  
TYPE: PRT  
ORGANISM: streptococcus mutans  
US-09-740-274-6

Query Match 59.8%: Score 67; DB 21; Length 1430;  
Best Local Similarity 59.1%: Pred. No. 0.091;  
Matches 13; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGARTINGQLYFRANGVQVG 22  
||:|||||:|||||:|||||  
DB 1332 TGSQTIAKKLYFASDGKQVKG 1353

RESULT 12  
US-09-499-203-2  
Sequence 2, Application US/09499203  
GENERAL INFORMATION:

Query Match	44.2%	Score 49.5;	DB 1;	length 811;
Best Local Similarity	52.2%;	Pred. No. 45;		
Matches 12;	Conservative 2;	Mismatches 8;	Indels 1;	Gaps 1;

Wed Mar 27 15:11:54 2002

us-09-290-049a-3.rapm

Page 6

```

0y      1 TGARTINGQLLYFRAN-GVQVKG 22
         |||||: || | ||
Db      587 TGLRTIDGKKYYFNTNTAVAVTG 609

```

Search completed: March 27, 2002, 14:20:24  
Job time: 1571 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:22:46 ; Search time 137.48 Seconds  
(without alignments)  
11.042 Million cell updates/sec

Title: US-09-290-049a-3  
Perfect score: 112  
Sequence: 1 TGAATGCOLLYFRANGVQVKG 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 332938 seqs, 68999538 residues

Total number of hits satisfying chosen parameters: 332938

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep1:\*  
7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep:\*  
8: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	112	100.0	22	6	US-09-562-328-37 Sequence 37, Appl
2	112	100.0	22	6	US-09-290-049a-3 Sequence 3, Appl
3	87	77.7	22	6	US-09-562-328-7 Sequence 7, Appl
4	87	77.7	26	6	US-09-562-328-43 Sequence 43, Appl
5	49.5	44.2	811	7	US-10-011-366-7 Sequence 7, Appl
6	49.5	44.2	2710	7	US-10-011-366-6 Sequence 6, Appl
7	45	40.2	563	6	US-09-708-427-10858 Sequence 10858, A
8	45	40.2	630	6	US-09-708-427-10857 Sequence 10857, A
9	44	39.3	115	7	US-10-091-504-970 Sequence 970, App
10	44	39.3	118	5	US-09-621-976-4566 Sequence 4566, App
11	44	39.3	1675	6	US-09-708-427-30688 Sequence 30688, A
12	44	39.3	1684	6	US-09-708-427-30687 Sequence 30687, A
13	44	39.3	1780	6	US-09-708-427-30686 Sequence 30686, A
14	43.5	38.8	152	6	US-09-605-703B-1602 Sequence 1602, Ap
15	43	38.4	706	6	US-09-848-909-26 Sequence 26, Appl
16	43	38.4	875	6	US-09-848-909-32 Sequence 32, Appl
17	42	37.5	944	6	US-09-897-516-5617 Sequence 5617, Ap
18	42	37.5	929	6	US-09-748-875-60 Sequence 60, Appl
19	41.5	37.1	472	7	US-10-029-386-33685 Sequence 33685, A
20	41	36.6	174	6	US-09-675-784A-7970 Sequence 7970, Ap
21	41	36.6	242	6	US-09-897-516-4915 Sequence 4915, Ap
22	41	36.6	383	7	US-10-015-127-9902 Sequence 9902, App
23	41	36.6	474	6	US-09-604-693A-106 Sequence 106, App
24	41	36.6	846	8	US-60-341-261-3302 Sequence 3302, Ap
25	40.5	36.2	438	7	US-10-015-127-13123 Sequence 13123, A

26	40	35.7	19	6	US-09-562-328-39 Sequence 39, Appl
27	40	35.7	169	6	US-09-605-703B-1180 Sequence 1180, Ap
28	40	35.7	231	6	US-09-631-616-17 Sequence 17, Appl
29	40	35.7	266	6	US-09-708-427-17269 Sequence 17269, A
30	40	35.7	363	6	US-09-708-427-17268 Sequence 17268, A
31	40	35.7	379	6	US-09-614-150-33408 Sequence 33408, A
32	40	35.7	400	6	US-09-605-703B-1178 Sequence 1178, Ap
33	40	35.7	441	6	US-09-708-427-17267 Sequence 17267, A
34	40	35.7	590	6	US-09-708-427-9652 Sequence 9652, Ap
35	40	35.7	638	6	US-09-708-427-9651 Sequence 9651, Ap
36	40	35.7	669	6	US-09-708-427-9651 Sequence 9651, Ap
37	39.5	35.3	304	7	US-10-015-127-11175 Sequence 11175, A
38	39	34.8	91	7	US-10-011-366-8 Sequence 8, Appl
39	39	34.8	137	6	US-09-614-150-15900 Sequence 15900, A
40	39	34.8	196	6	US-09-708-427-9649 Sequence 9649, Ap
41	39	34.8	219	6	US-09-708-427-9648 Sequence 9648, Ap
42	39	34.8	319	6	US-09-708-427-9645 Sequence 9645, Ap
43	39	34.8	351	6	US-09-675-784A-8859 Sequence 8859, Ap
44	39	34.8	371	6	US-09-708-427-9647 Sequence 9647, Ap
45	39	34.8	393	7	US-10-015-127-12338 Sequence 12338, A

## ALIGNMENTS

```
RESULT 1
US-09-562-328-37
: Sequence 37, Application US/09562328
: GENERAL INFORMATION:
: APPLICANT: LEEB, ANDREW
: APPLICANT: TAUBMAN, MARTIN A.
: TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
: FILE REFERENCE: 04995.0046-01
: CURRENT APPLICATION NUMBER: US/09/562.328
: PRIOR FILING DATE: 2000-05-01
: PRIOR APPLICATION NUMBER: 09/288,965
: PRIOR FILING DATE: 1999-04-09
: NUMBER OF SEQ ID NOS: 44
: SOFTWARE: Patentn Ver. 2.1
: SEQ ID NO 37
: LENGTH: 22
: TYPE: PRT
: ORGANISM: Streptococcus sp.
US-09-562-328-37

Query Match      100.0%; Score 112; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.8e-12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGAATGCOLLYFRANGVQVKG 22
      |||||
Db      1 TGAATGCOLLYFRANGVQVKG 22

RESULT 2
US-09-290-049a-3
: Sequence 3, Application US/09290049A
: GENERAL INFORMATION:
: APPLICANT: Smith, Daniel J.
: APPLICANT: Taubman, Martin A.
: TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
: FILE REFERENCE: 1564.1008-002
: CURRENT APPLICATION NUMBER: US/09/290.049A
: PRIOR FILING DATE: 1999-04-12
: PRIOR APPLICATION NUMBER: 60/081,550
: PRIOR FILING DATE: 1998-04-13
: PRIOR APPLICATION NUMBER: 60/115,142
: PRIOR FILING DATE: 1999-01-08
: NUMBER OF SEQ ID NOS: 19
: SOFTWARE: FastSeq for Windows Version 4.0
```

Query Match	77.7%;	Score 87;	DB 6;	Length 26
*Best Logical Similarity	77.3%;	Pred. No. 5.3e-08;		

RESULT 6  
US-10-011-366-6

; Sequence 6, Application US/10011366  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES  
; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE  
; DISEASE  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/011,366  
; FILING DATE: 16-Nov-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/957,310  
; FILING DATE: 23-OCT-1997  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 24-OCT-1994  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPND-01121  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2710 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-10-011-366-6  
Query Match 44.2%; Score 49.5; DB 7; Length 2710;  
Best Local Similarity 52.2%; Pred. No. 25;  
Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;  
OY 1 TGARTINGOLLYFRAN-GVQYKG 22  
DB 2460 TGLRTIDGKKYFNTNTAVATG 2482  
RESULT 7  
US-09-708-427-10858  
; Sequence 10858, Application US/09708427  
; GENERAL INFORMATION:  
; APPLICANT: N. Alexandrov et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1243P  
; CURRENT APPLICATION NUMBER: US/09/708,427  
; CURRENT FILING DATE: 2000-11-09  
; NUMBER OF SEQ ID NOS: 85364  
; SOFTWARE: Patentin version 3.1

; SEQ ID NO 10858  
; LENGTH: 563  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc-feature  
; LOCATION: 1..563  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc-feature  
; LOCATION: 1..563  
; OTHER INFORMATION: Ceres Seq. ID 1822113  
US-09-708-427-10858  
Query Match 40.2%; Score 45; DB 6; Length 563;  
Best Local Similarity 45.5%; Pred. No. 22;  
Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;  
OY 1 TGARTINGOLLYFRANGVQYKG 22  
DB 72 TGIGPIHGRICMFMANDPYVG 93  
RESULT 8  
US-09-708-427-10857  
; Sequence 10857, Application US/09708427  
; GENERAL INFORMATION:  
; APPLICANT: N. Alexandrov et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID  
; FILE REFERENCE: 2750-1243P  
; CURRENT APPLICATION NUMBER: US/09/708,427  
; CURRENT FILING DATE: 2000-11-09  
; NUMBER OF SEQ ID NOS: 85364  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 10857  
; LENGTH: 630  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc-feature  
; LOCATION: 1..630  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc-feature  
; LOCATION: 1..630  
; OTHER INFORMATION: Ceres Seq. ID 1822112  
US-09-708-427-10857  
Query Match 40.2%; Score 45; DB 6; Length 630;  
Best Local Similarity 45.5%; Pred. No. 26;  
Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;  
OY 1 TGARTINGOLLYFRANGVQYKG 22  
DB 139 TGIGPIHGRICMFMANDPYVG 160  
RESULT 9  
US-10-091-504-970  
; Sequence 970, Application US/10091504  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC007C1  
; CURRENT APPLICATION NUMBER: US/10/091,504  
; CURRENT FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 2442  
; Prior Application removed - See File Wrapper or Palm  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 970  
; LENGTH: 115  
; TYPE: PRT

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (26)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (28)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (30)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (46)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (50)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (58)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (59)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-091-504-970
```

```

Query Match          39.3%; Score 44; DB 7; Length 115;
Best Local Similarity 70.0%; Pred. No. 5.1;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

5 TINGOLLYR 14
I:||||:|:|
63 TVNGELLFFR 72
```

```

RESULT 10
US-09-621-976-4566
; Sequence 4566, Application US/09621976
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 4566
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-4566
```

```

Query Match          39.3%; Score 44; DB 5; Length 148;
Best Local Similarity 70.0%; Pred. No. 6.8;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      5 TINGOLLYR 14
I:||||:|:|
DB      63 TVNGELLFFR 72
```

```

RESULT 11
US-09-708-427-30688
; Sequence 30688, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30688
; LENGTH: 1675
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..1675
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..1675
; OTHER INFORMATION: Ceres Seq. ID 1830036
US-09-708-427-30688
```

```

Query Match          39.3%; Score 44; DB 6; Length 1675;
Best Local Similarity 53.3%; Pred. No. 1.2e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY      1 TCGARTINGOLLYERA 15
I||||:|:|:|
DB      881 TLAATVRCGMYYRRA 895
```

```

RESULT 12
US-09-708-427-30687
; Sequence 30687, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30687
; LENGTH: 1684
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..1684
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..1684
; OTHER INFORMATION: Ceres Seq. ID 1830035
US-09-708-427-30687
```

```

Query Match          39.3%; Score 44; DB 6; Length 1684;
Best Local Similarity 53.3%; Pred. No. 1.2e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY      1 TCGARTINGOLLYERA 15
I||||:|:|:|
```



Db 890 TLARTVRCGMYYRA 904

## RESULT 13

US-09-708-427-30686  
; Sequence 30686, Application US/09708427  
; GENERAL INFORMATION:  
; APPLICANT: N. ALEXANDROV et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1243P  
; CURRENT APPLICATION NUMBER: US/09/708.427  
; PRIOR FILING DATE: 2000-11-09  
; NUMBER OF SEQ ID NOS: 85364  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 30686  
; LENGTH: 1780  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: 1..1780  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc.feature  
; LOCATION: 1..1780  
; OTHER INFORMATION: Ceres Seq. ID 1830034  
US-09-708-427-30686

## Query Match

Best Local Similarity 39.3%; Score 44; DB 6; Length 1780;  
Best Local Similarity 53.3%; Pred. No. 1.3e+02;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGARTINGQLYFRA 15

Db 986 TLARTVRCGMYYRA 1000

## RESULT 14

US-09-605-703B-1602  
; Sequence 1602, Application US/09605703B  
; GENERAL INFORMATION:  
; APPLICANT: Pompeius, Markus  
; APPLICANT: Krieger, Burkhard  
; APPLICANT: Schroder, Hartwig  
; APPLICANT: Zelder, Oskar  
; APPLICANT: Haberhauser, Gregor  
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL  
; FILE REFERENCE: BGI-129CP  
; CURRENT APPLICATION NUMBER: US/09/605.703B  
; PRIOR FILING DATE: 2000-06-27  
; PRIOR APPLICATION NUMBER: 60/142,764  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: 60/152,318  
; PRIOR FILING DATE: 1999-09-03  
; NUMBER OF SEQ ID NOS: 2934  
; SEQ ID NO 1602  
; LENGTH: 152  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-605-703B-1602

## Query Match

Best Local Similarity 38.8%; Score 43.5; DB 6; Length 152;  
Best Local Similarity 54.5%; Pred. No. 8.5;  
Matches 12; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 1 TGARTINGQLYFRANGVQYK 22

Db 125 TGA-TISASANYLRAAGVVRG 145

## RESULT 15

US-09-848-909-26  
; Sequence 26, Application US/09848909  
; GENERAL INFORMATION:  
; APPLICANT: Collier, R. John  
; APPLICANT: Sellman, Brett R.  
; TITLE OF INVENTION: Compounds and Methods for the Treatment  
; FILE REFERENCE: 00742/060002  
; CURRENT APPLICATION NUMBER: US/09/848.909  
; PRIOR FILING DATE: 2001-05-04  
; PRIOR APPLICATION NUMBER: US 60/201,800  
; PRIOR FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PastSeq for Windows Version 4.0  
; SEQ ID NO 26  
; LENGTH: 706  
; TYPE: PRT  
; ORGANISM: Clostridium perfringens  
US-09-848-909-26

## Query Match

Best Local Similarity 38.4%; Score 43; DB 6; Length 706;  
Best Local Similarity 56.2%; Pred. No. 63;  
Matches 9; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 5 TINGQLYFRANGVQY 20

Db 388 TKNGLYF--NGIRP 401

Search completed: March 27, 2002, 14:22:46  
Job time: 1693 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:01:16 ; Search time 102.51 Seconds  
(without alignments)  
16.348 Million cell updates/sec

Title: US-09-290-049a-3  
Perfect score: 112  
Sequence: 1 TCGARTINGQLTYFRANGVQVKG 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: PIR\_68:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112	100.0	1475	2 B33135	glfB protein precu
2	87	77.7	1592	2 A38175	glucosyltransferas
3	82	73.2	1599	2 S22737	glucosyltransferas
4	80	71.4	1518	2 A44811	glucosyltransferas
5	78	69.6	1375	2 J70345	dextranucrase (EC
6	75	67.0	1290	2 J05473	dextranucrase (EC
7	75	67.0	1365	2 A41483	glucosyltransferas
8	73	65.2	1449	2 T30857	glucosyltransferas
9	72	64.3	1508	2 T31098	probable dextranu
10	68	60.7	1449	2 A37184	glucan-binding pro
11	68	60.7	1449	2 T30552	glucosyltransferas
12	67	59.8	1431	2 A45866	dextranucrase (EC
13	66	58.9	1577	2 T30858	dextranucrase (EC
14	65.5	58.5	329	2 A55221	dextranucrase (EC
15	54	48.2	221	2 T37244	GATA transcription
16	50	44.6	831	2 T00323	chitinase (EC 3.2.
17	49.5	44.2	529	2 T10388	hypothetical prote
18	49.5	44.2	2710	2 A37052	hypothetical prote
19	47	42.0	126	2 S40016	toxin A - Clostrid
20	47	42.0	696	2 A29635	doc protein - phag
21	47	42.0	788	2 J50747	transcription fact
22	46	41.1	231	2 F70471	conserved hypotet
23	46	41.1	245	2 F64465	hypothetical prote
24	46	41.1	301	2 S13165	asialoglycoprotein
25	46	41.1	619	2 A41971	surface protein ps
26	46	41.1	1301	2 S18118	alpha-amylase - Al
27	46	41.1	2364	2 I40884	cytotoxin L - Clo
28	45	40.2	630	2 T05433	hypothetical prote
29	45	40.2	648	2 S10869	enterotoxin A - Cl

30	44	39.3	329	2 F71014	probable dipeptide
31	44	39.3	522	2 D71074	probable methylal
32	44	39.3	522	2 H75135	methylmalonyl-coa
33	44	39.3	566	2 S74633	high affinity sulf
34	44	39.3	591	2 C84220	propionyl-CoA carb
35	44	39.3	829	2 T29372	hypothetical prote
36	44	39.3	1396	2 S36851	L-shaped tail fibe
37	44	39.3	1679	2 T30271	surface protein -
38	44	39.3	1780	2 A85045	probable glucan sy
39	44	39.3	2178	2 S5805	alpha-toxin - Clo
40	43	38.4	110	2 B69403	hypothetical prote
41	43	38.4	299	2 T15928	hypothetical prote
42	43	38.4	305	2 B75474	probable chloride
43	43	38.4	329	2 H75117	peptide abc transp
44	43	38.4	467	1 S45493	serine proteinase
45	43	38.4	509	2 A49094	methylmalonyl-CoA

## ALIGNMENTS

```
RESULT 1
B33135
glfB protein precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 15-Oct-1999
C:Accession: B33135, A33128
R:Shiroza, T.; Ueda, S.; Kuramitsu, H. K.
J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the glfB gene from Streptococcus mutans.
A:Reference number: A33135; MUID:87308013
A:Accession: B33135
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1475 <SH2>
A:Cross-references: GB:M17361; NID:G153639; PIDN:AAA8588.1; PID:G153640
R:Shiroza, T.; Ueda, S.; Kuramitsu, H. K.
submitted to the Protein Sequence Database, September 1990
A:Reference number: A33128
A:Accession: A33128
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-171, 173-641, 'N', 643-1475 <SH2>
A:Experimental source: strain GS-5
C:Superfamily: cpl repeat homology
F:1096-1115/Domain: cpl repeat homology <CP1>
F:1124-1243/Domain: cpl repeat homology <CP2>
F:1289-1308/Domain: cpl repeat homology <CP3>
F:1354-1373/Domain: cpl repeat homology <CP4>
F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match 100.0%; Score 112; DB 2; Length 1475;
Best Local Similarity 100.0%; Pred. No. 4.9e-10;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGARTINGQLTYFRANGVQVKG 22
Db 1300 TCGARTINGQLTYFRANGVQVKG 1321

RESULT 2
A38175
glucosyltransferase precursor - Streptococcus sobrius
C:Species: Streptococcus sobrius
C:Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 15-Oct-1999
C:Accession: A38175
R:Abu, H.; Msumura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.
J. Bacteriol. 173, 989-996, 1991
A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococ
A:Reference number: A38175; MUID:91123227
A:Accession: A38175
A>Status: preliminary
```

A:Molecule type: DNA  
A:Residues: 1-1592 <ABO>  
A:Cross-references: GB:D90213; NID:g217032; PIDN:BAAL4241.1; PID:01014946; PID:g217033  
C:Superfamily: cpl repeat homology  
F:1093-1112/Domain: cpl repeat homology <CP1>  
F:1222-1241/Domain: cpl repeat homology <CP2>  
F:1287-1306/Domain: cpl repeat homology <CP3>  
F:1330-1351/Domain: cpl repeat homology <CP4>  
F:1352-1371/Domain: cpl repeat homology <CP5>  
F:1402-1420/Domain: cpl repeat homology <CP6>  
F:1465-1484/Domain: cpl repeat homology <CP7>  
F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 77.7%; Score 87; DB 2; Length 1592;  
Best Local Similarity 77.3%; Pred. No. 7e-06;  
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

1 TGARTINGOLLYFRANGOVKG 22  
||||| || ||||| |||||  
Db 1298 TGAQRTGCKLYFRANGOVKG 1319

RESULT 3  
S22737  
glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000  
C:Accession: S22737; S28810; B44811; S22727  
R:Jacques, N.  
A:Reference number: S22726  
A:Accession: S22737  
A:Molecule type: DNA  
A:Residues: 1-1599 <JAC>  
A:Cross-references: EMBL:211872; NID:g47530; PIDN:CAA77898.1; PID:g47531  
A:Experimental source: ATCC 25975  
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
J. Gen. Microbiol. 137, 2577-2593, 1991  
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase gen  
A:Reference number: A44811; MUID:92148377  
A:Accession: S28810  
A:Molecule type: DNA  
A:Residues: 1-51 <GIF>  
A:Cross-references: EMBL:211873  
A:Genetics:  
A:Gene: gltK  
C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:1456-1475/Domain: cpl repeat homology <CPR>

Query Match 73.2%; Score 82; DB 2; Length 1599;  
Best Local Similarity 77.3%; Pred. No. 4.7e-05;  
Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGARTINGOLLYFRANGOVKG 22  
||||| ||||| ||||| |||||  
Db 1402 TGAQYINGOLLYFRANGOVKG 1423

RESULT 4  
A44811  
glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 15-Oct-1999  
C:Accession: A44811; S22726; S28809  
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
J. Gen. Microbiol. 137, 2577-2593, 1991  
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase gen  
A:Reference number: A44811; MUID:92148377  
A:Accession: A44811  
A:Molecule type: DNA

A:Residues: 1-1518 <GIF>  
A:Cross-references: EMBL:211873; NID:g47526; PIDN:CAA77900.1; PID:g47527  
A:Note: sequence extracted from NCBI backbone (NCBIN:81050, NCBI:81052)  
C:Genetics:  
A:Gene: gltK  
C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:1307-1326/Domain: cpl repeat homology <CP4>

Query Match 71.4%; Score 80; DB 2; Length 1518;  
Best Local Similarity 72.7%; Pred. No. 9.5e-05;  
Matches 16; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGARTINGOLLYFRANGOVKG 22  
||||| ||||| ||||| |||||  
Db 1383 TGAQYINGOLLYFRANGOVKG 1404

RESULT 5  
JT0345  
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)  
N:Alternate names: sucrose 6-glucosyltransferase  
C:Species: Streptococcus mutans  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 24-Sep-1999  
R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.  
Gene 69, 101-109, 1988  
A:Title: Sequence analysis of the gltC gene from Streptococcus mutans GS-5.  
A:Reference number: JT0345; MUID:89137980  
A:Accession: JT0345  
A:Molecule type: DNA  
A:Residues: 1-1375 <UED>  
A:Experimental source: GS-5  
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
J. Bacteriol. 169, 4263-4270, 1987  
A:Title: Sequence analysis of the gltB gene from Streptococcus mutans.  
A:Reference number: A33135; MUID:87308013  
A:Accession: C33135  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-349 <SHI>  
A:Cross-references: GB:M17361  
C:Genetics:  
A:Gene: gltC  
C:Function:  
A:Description: catalyzes the synthesis of both water-soluble and water-insoluble gluc  
C:Superfamily: cpl repeat homology  
C:Keywords: duplication; glycosyltransferase; hexosyltransferase  
F:1-34/Domain: signal sequence #status predicted <SIG>  
F:35-1375/Product: glucosyltransferase #status predicted <MAT>  
F:1126-1145/Domain: cpl repeat homology <CP1>  
F:1125-1172/Domain: cpl repeat homology <CP2>  
F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 69.6%; Score 78; DB 2; Length 1375;  
Best Local Similarity 68.2%; Pred. No. 0.00018;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGARTINGOLLYFRANGOVKG 22  
||||| ||||| ||||| |||||  
Db 1264 TGTVTFTNGCKLYFRANGOVKG 1285

RESULT 6  
JC5473  
dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides  
C:Species: Leuconostoc mesenteroides  
C:Date: 07-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 29-Aug-1997  
C:Accession: JC5473  
R:Monchois, V.; Willems, R.M.; Renaud-Simeon, M.; Croux, C.; Monsan, P.  
Gene 182, 23-32, 1996

A:Title: Cloning and sequencing of a gene coding for a novel dextranucrase from *Leuconostoc mesenteroides*  
A:Reference number: J05473; MUID:9713686  
A:Accession: J05473  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1290 <MON>  
A:Cross-references: GB:058181  
C:Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose onto  
C:Genetics:  
A:Gene: dsrA  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:78-870/Domain: catalytic #status predicted <CAT>  
F:922-1290/Domain: glucan-binding #status predicted <GCB>

Query Match 67.0%; Score 75; DB 2; Length 1290;  
Best Local Similarity 68.2%; Pred. No. 0.00053;  
Matches 15; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 TGARTINGOLLYFRANGOVKG 22  
||:|||||||:|||||  
P 1245 TGVVINGOLLYFDAGROVKG 1266

RESULT 7  
A1483  
glucosyltransferase (EC 2.4.1.-) gtfS precursor - *Streptococcus sobrinus*  
C:Species: *Streptococcus sobrinus*  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999  
C:Accession: A1483  
R:Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.  
Infect. Immun. 58, 2452-2458, 1990  
A:Title: Analysis of the *Streptococcus downei* gtfS gene, which specifies a glucosyltrans  
A:Reference number: A1483; MUID:9031665  
A:Accession: A1483  
A:Molecule type: DNA  
A:Residues: 1-1365 <GTL>  
A:Cross-references: GB:M30943; NID:g153652; PIDN:AAA26898.1; PID:g153653  
C:Genetics:  
A:Gene: gtfS  
C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 67.0%; Score 75; DB 2; Length 1365;  
Best Local Similarity 63.6%; Pred. No. 0.00057;  
Matches 14; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGARTINGOLLYFRANGOVKG 22  
||:|||||||:|||||  
P 1235 TGEQTIDGKVFQNGOVKG 1256

RESULT 8  
T30857  
glucosyltransferase - *Streptococcus salivarius*  
C:Species: *Streptococcus salivarius*  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30857  
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
Infect. Immun. 63, 609-621, 1995  
A:Title: *Streptococcus salivarius* ATCC 25975 possesses at least two genes coding for pri  
A:Reference number: Z20909; MUID:95122197  
A:Accession: T30857  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1449 <STM>  
A:Cross-references: EMBL:L35495; NID:g662378; PID:g662379; PIDN:AAC41412.1  
C:Genetics:  
A:Gene: gtfL

Query Match 65.2%; Score 73; DB 2; Length 1449;

Best Local Similarity 68.2%; Pred. No. 0.0013;  
Matches 15; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 TGARTINGOLLYFRANGOVKG 22  
||:|||||||:|||||  
DB 1258 TGHONINGELFEDNNGOVKG 1279

RESULT 9  
T31098  
probable dextranucrase (EC 2.4.1.5), extracellular - *Leuconostoc mesenteroides*  
C:Species: *Leuconostoc mesenteroides*  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
C:Accession: T31098  
R:Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willemot, R.M.  
FEBS Microbiol. Lett. 159, 307-315, 1998  
A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase  
A:Reference number: Z20981; MUID:98164374  
A:Accession: T31098  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1508 <MON>  
A:Cross-references: EMBL:AF030129; NID:g2766611; PID:g2766612; PIDN:AAB95453.1  
A:Experimental source: strain NRRL B-1299  
C:Genetics:  
A:Gene: dsrB  
C:Function:  
A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 64.3%; Score 72; DB 2; Length 1508;  
Best Local Similarity 68.2%; Pred. No. 0.002;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 TGARTINGOLLYFRANGOVKG 22  
||:|||||||:|||||  
DB 1399 TGLVINGOLLYFQANGROVKG 1420

RESULT 10  
A37184  
glucan-binding protein - *Streptococcus mutans*  
C:Species: *Streptococcus mutans*  
C:Date: 31-Jan-1992 #sequence\_revision 31-Jan-1992 #text\_change 15-Oct-1999  
C:Accession: A37184  
R:Banas, J.A.; Russell, R.R.B.; Ferretti, J.J.  
Infect. Immun. 58, 667-673, 1990  
A:Title: Sequence analysis of the gene for the glucan-binding protein of *Streptococu*  
A:Reference number: A37184; MUID:90170123  
A:Accession: A37184  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-561 <TAN>  
A:Cross-references: GB:M30945; NID:g153637; PIDN:AAA26894.1; PID:g153638  
C:Superfamily: cpl repeat homology  
F:169-188/Domain: cpl repeat homology <CP1>  
F:264-283/Domain: cpl repeat homology <CP2>  
F:349-368/Domain: cpl repeat homology <CP3>  
F:504-523/Domain: cpl repeat homology <CP4>  
F:525-548/Domain: cpl repeat homology <CP5>

Query Match 60.7%; Score 68; DB 2; Length 563;  
Best Local Similarity 66.7%; Pred. No. 0.0031;  
Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 2 GARTINGOLLYFRANGOVKG 22  
| ||| | : || ||||| |  
DB 276 GMRITGGKRYVFPDNGOVKG 296

RESULT 11

T30552  
glucosyltransferase N - Streptococcus salivarius (fragment)  
A:Accession: T30552  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1449 <JAF>  
A:Cross-references: EMBL:AF049609; NID:g2935545; PID:g2935546; PIDN:AAC05156.1  
C:Genetics:  
A:Gene: gln

Submitted to the EMBL Data Library, February 1998  
A:Description: Streptococcus salivarius V1477 gln.  
A:Reference number: 220854  
A:Accession: T30552  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1449 <JAF>  
A:Cross-references: EMBL:AF049609; NID:g2935545; PID:g2935546; PIDN:AAC05156.1  
C:Genetics:  
A:Gene: gln

Query Match 60.7%; Score 68; DB 2; Length 1449;  
Best Local Similarity 66.7%; Pred. No. 0.0086;  
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 GARTINGQLLYFRANGVOYKG 22  
1 : ||||| : ||||| |||||  
DB 1259 GHQINGQELFDNGVOYKG 1279

RESULT 12

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans  
A:Accession: A45866  
C:Species: Streptococcus mutans  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: A45866  
R:Honda, O.; Kato, C.; Kuramitsu, H.K.  
J. Gen. Microbiol. 136, 2099-2105, 1990  
A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl  
A:Reference number: A45866; MUID:91100958  
A:Accession: A45866  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1431 <HON>  
A:Cross-references: GB:M29296  
C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:181-201/Domain: cpl repeat homology <CP1>  
F:127-1146/Domain: cpl repeat homology <CP2>  
F:92-1211/Domain: cpl repeat homology <CP3>  
F:1257-1276/Domain: cpl repeat homology <CP4>  
F:1277-1297/Domain: cpl repeat homology <CP5>  
F:1321-1340/Domain: cpl repeat homology <CP8>  
F:1341-1361/Domain: cpl repeat homology <CP6>  
F:1365-1404/Domain: cpl repeat homology <CP7>

Query Match 59.8%; Score 67; DB 2; Length 1431;  
Best Local Similarity 59.1%; Pred. No. 0.012;  
Matches 13; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 1 TGARTINGQLLYFRANGVOYKG 22  
1 : ||||| : ||||| : |||||  
DB 1332 TGSQTAGKLYFASDQKVOYKG 1353

RESULT 13

glucosyltransferase - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30858  
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
Infect. Immun. 63, 609-621, 1995  
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pr  
A:Reference number: 220909; MUID:95122197

A:Accession: T30858  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1577 <SIM>  
A:Cross-references: EMBL:L35928; NID:g662380; PID:g662381; PIDN:AAC41413.1  
C:Genetics:  
A:Gene: gtfm

Query Match 58.9%; Score 66; DB 2; Length 1577;  
Best Local Similarity 54.5%; Pred. No. 0.02;  
Matches 12; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 1 TGARTINGQLLYFRANGVOYKG 22  
1 : ||||| : ||||| : |||||  
DB 1469 TGLQININRKYFGSNGAOYKG 1490

RESULT 14

dextranase inhibitor precursor - Streptococcus sobrinus (strain UAB66, serotype g)  
A:Accession: A55221  
C:Species: Streptococcus sobrinus  
C:Date: 23-Mar-1995 #sequence\_revision 23-Mar-1995 #text\_change 17-Mar-1999  
C:Accession: A55221  
R:Sun, J.W.; Wanda, S.Y.; Camilli, A.; Curtiss III, R.  
J. Bacteriol. 176, 7213-7222, 1994  
A:Title: Cloning and DNA sequencing of the dextranase inhibitor gene (dei) from Strep  
A:Reference number: A55221; MUID:95050304  
A:Accession: A55221  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-329 <SUN>  
A:Cross-references: GB:L34406  
C:Superfamily: cpl repeat homology  
F:173-192/Domain: cpl repeat homology <CP1>

Query Match 58.5%; Score 65.5; DB 2; Length 329;  
Best Local Similarity 68.2%; Pred. No. 0.0044;  
Matches 15; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

OY 2 GARTINGQLLYFRAN-GVOYKG 22  
1 : ||||| : ||||| : |||||  
DB 185 GLQTINGQLLYFDTNKQVOYKG 206

RESULT 15

GATA transcription factor end-1 - Caenorhabditis elegans  
T37244  
C:Species: Caenorhabditis elegans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jul-2000  
C:Accession: T37244; T22912  
R:Zhu, J.; Hill, R.J.; Held, P.J.; Fukuyama, M.; Sugimoto, A.; Priest, J.R.; Rothman,  
Genes Dev. 11, 2883-2896, 1997  
A:Title: End-1 encodes an apparent GATA factor that specifies the endoderm precursor  
A:Reference number: 221642; MUID:98019168  
A:Accession: T37244  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-221 <ZHU>  
A:Cross-references: EMBL:AF026555; NID:g2655205; PIDN:AAB97516.1; PID:g2655206  
R:McMurray, A.  
Submitted to the EMBL Data Library, November 1996  
A:Reference number: 219637  
A:Accession: T22912  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-221 <WILL>  
A:Cross-references: EMBL:Z81555; PIDN:CAB04513.1; GSPDB:GND0023; CESP:F58E10.2  
A:Experimental source: clone F58E10  
C:Genetics:  
A:Gene: end-1; F58E10.2  
A:Map position: 5

A: Introns: 71/1; 128/3; 162/2

C: Function:

A: Description: involved in the specification of germ layer and founder cell identity

C: Keywords: transcription factor; zinc finger

Query Match

48.2%; Score 54; DB 2; Length 221;

Best Local Similarity 57.9%; Pred. No. 0.23;

Matches 11; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 1 TGARTINGQLYFRANGVQ 19

: 11 11 11 11 11 11

Db 156 SGAIECNCGLYFRKNGIQ 174

Search completed: March 27, 2002, 14:01:17  
JOD time: 479 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:54:13 ; Search time 1139.61 seconds  
(without alignments)  
5.116 Million cell updates/sec

Title: US-09-290-049a-1

Sequence: 1 ANDHLSILEAWSNDPTPLD 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

Number of hits satisfying chosen parameters: 3148936

Maximum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/2/paa/PCRTUS\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US080\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep:\*  
8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep:\*  
9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep:\*  
10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep:\*  
11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep:\*  
12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep:\*  
13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep:\*  
14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep:\*  
15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep:\*  
16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep:\*  
17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep:\*  
18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep:\*  
19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep:\*  
20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep:\*  
21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep:\*  
22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep:\*  
23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep:\*  
24: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	21	16	US-09-290-049-1
2	119	100.0	21	16	US-09-290-049-10
3	119	100.0	1475	19	US-09-557-848-2
4	119	100.0	1475	21	US-09-740-274-2
5	110	90.8	1375	21	US-09-740-274-4
6	108	90.8	21	16	US-09-290-049-12
7	108	90.8	21	16	US-09-290-049-14
8	75	63.0	21	16	US-09-290-049-11
9	75	63.0	1430	20	US-09-649-885-2

10	75	63.0	1430	21	US-09-740-274-6	Sequence 6, Appl1
11	73	61.3	21	16	US-09-290-049-13	Sequence 13, Appl1
12	60	50.4	2057	18	US-09-499-203-2	Sequence 2, Appl1
13	50	42.0	306	19	US-09-595-298A-661	Sequence 661, App
14	50	42.0	306	19	US-09-595-298A-660	Sequence 660, App
15	49.5	41.6	546	24	US-60-324-109-17015	Sequence 17015, A
16	49.5	41.6	549	24	US-60-324-109-16789	Sequence 16789, A
17	49.5	41.2	391	24	US-60-324-109-20736	Sequence 20736, A
18	48.5	40.8	401	1	PCT-US01-01309-222	Sequence 222, App
19	48	40.3	773	16	US-09-248-796-22573	Sequence 22573, A
20	47.5	39.9	178	22	US-09-867-716-19848	Sequence 19848, A
21	47.5	39.9	534	17	US-09-344-882-24	Sequence 24, Appl1
22	47.5	39.9	534	24	US-60-324-109-33149	Sequence 33149, A
23	47	39.5	196	19	US-09-595-298A-765	Sequence 765, App
24	46.5	39.1	220	24	US-60-212-413-164	Sequence 164, App
25	46.5	39.1	220	24	US-60-229-518-331	Sequence 331, App
26	46.5	39.1	330	24	US-60-207-422-75	Sequence 75, Appl1
27	46.5	39.1	539	1	PCT-US01-01332-762	Sequence 762, App
28	46.5	39.1	539	21	US-09-764-875-762	Sequence 26597, A
29	46	38.7	62	18	US-09-417-507-26597	Sequence 5595, Ap
30	46	38.7	94	1	PCT-US01-08656-5595	Sequence 221, App
31	46	38.7	118	20	US-09-673-958A-221	Sequence 4571, Ap
32	46	38.7	199	21	US-09-738-626-4571	Sequence 348, App
33	46	38.7	275	19	US-09-538-092-348	Sequence 662, App
34	46	38.7	336	15	US-09-198-452A-662	Sequence 626, App
35	46	38.7	336	18	US-09-438-185-626	Sequence 8822, Ap
36	46	38.7	394	21	US-09-739-449-8822	Sequence 8822, Ap
37	46	38.7	394	22	US-09-803-110-8822	Sequence 9, Appl1
38	46	38.7	476	1	PCT-US01-122636-9	Sequence 9, Appl1
39	46	38.7	476	24	US-60-219-231-9	Sequence 21255, A
40	46	38.7	596	16	US-09-252-991A-21255	Sequence 7203, Ap
41	46	38.7	2590	24	US-60-215-161-7203	Sequence 5891, Ap
42	46	38.7	2603	24	US-60-215-161-5891	Sequence 5719, Ap
43	45	37.8	185	24	US-60-215-161-5719	Sequence 9825, Ap
44	45	37.8	194	1	PCT-US01-14827-9825	Sequence 80, Appl1
45	45	37.8	236	15	US-09-154-750-80	

#### ALIGNMENTS

RESULT 1  
US-09-290-049-1  
Sequence 1, Application US/09290049  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
FILE REFERENCE: PDC98-01P2A  
CURRENT APPLICATION NUMBER: US/09/290, 049  
EARLIER FILING DATE: 1999-04-12  
EARLIER APPLICATION NUMBER: 60/081, 550  
EARLIER FILING DATE: 1998-04-13  
EARLIER APPLICATION NUMBER: 60/115, 142  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 21  
TYPE: PPT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: EAW peptide  
US-09-290-049-1

Query Match 100.0%; Score 119; DB 16; Length 21;  
Best Local Similarity 100.0%; Pred. No. 5.3e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ANDHLSILEAWSNDPTPLD 21  
|||||

Db 1 ANDHLSILEAWSNDNTPYLHD 21

## RESULT 2

US-09-290-049-10  
; Sequence 10, Application US/09290049  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; FILE REFERENCE: FDC98-0192A  
; CURRENT APPLICATION NUMBER: US/09/290,049  
; CURRENT FILING DATE: 1999-04-12  
; EARLIER APPLICATION NUMBER: 60/081,550  
; EARLIER FILING DATE: 1998-04-13  
; EARLIER APPLICATION NUMBER: 60/115,142  
; EARLIER FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 10  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: S. mutans  
US-09-290-049-10

Query Match 100.0%; Score 119; DB 16; Length 21;  
Best Local Similarity 100.0%; Pred. No. 5.3e-11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDNTPYLHD 21  
Db 1 ANDHLSILEAWSNDNTPYLHD 21

## RESULT 3

US-09-557-848-2  
; Sequence 2, Application US/09557848  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; FILE REFERENCE: 0356D2  
; CURRENT APPLICATION NUMBER: US/09/557,848  
; CURRENT FILING DATE: 2000-04-26  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-557-848-2

Query Match 100.0%; Score 119; DB 19; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 5.2e-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDNTPYLHD 21  
Db 481 ANDHLSILEAWSNDNTPYLHD 501

## RESULT 4

US-09-740-274-2  
; Sequence 2, Application US/09740274  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.

; TITLE OF INVENTION: Glucan-containing Compositions and Paper

; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-740-274-2

Query Match 100.0%; Score 119; DB 21; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 5.2e-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDNTPYLHD 21  
Db 481 ANDHLSILEAWSNDNTPYLHD 501

## RESULT 5

US-09-740-274-4  
; Sequence 4, Application US/09740274  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-740-274-4

Query Match 92.4%; Score 110; DB 21; Length 1375;  
Best Local Similarity 95.2%; Pred. No. 1.2e-07;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDNTPYLHD 21

Db 507 ANDHLSILEAMSDNDTPYLHD 527

RESULT 6

US-09-290-049-12

; Sequence 12, Application US/09290049

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.

; APPLICANT: Taubman, Martin A.

; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL

; FILE REFERENCE: FDC98-01P2A

; CURRENT APPLICATION NUMBER: US/09/290,049

; CURRENT FILING DATE: 1999-04-12

; EARLIER APPLICATION NUMBER: 60/081,550

; EARLIER FILING DATE: 1998-04-13

; EARLIER APPLICATION NUMBER: 60/115,142

; EARLIER FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 12

; LENGTH: 21

; TYPE: PRT

; ORGANISM: S. downei

US-09-290-049-12

Query Match

Best Local Similarity 90.8%; Score 108; DB 16; Length 21;

Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDTPYLHD 21

Db 1 ANNHVSIVEAMSDNDTPYLHD 21

RESULT 7

US-09-290-049-14

; Sequence 14, Application US/09290049

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.

; APPLICANT: Taubman, Martin A.

; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL

; FILE REFERENCE: FDC98-01P2A

; CURRENT APPLICATION NUMBER: US/09/290,049

; CURRENT FILING DATE: 1999-04-12

; EARLIER APPLICATION NUMBER: 60/081,550

; EARLIER FILING DATE: 1998-04-13

; EARLIER APPLICATION NUMBER: 60/115,142

; EARLIER FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 14

; LENGTH: 21

; TYPE: PRT

; ORGANISM: S. sobrinus

US-09-290-049-14

Query Match

Best Local Similarity 90.8%; Score 108; DB 16; Length 21;

Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDTPYLHD 21

Db 1 ANNHVSIVEAMSDNDTPYLHD 21

RESULT 8

US-09-290-049-11

; Sequence 11, Application US/09290049

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.

; APPLICANT: Taubman, Martin A.

; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL

; FILE REFERENCE: FDC98-01P2A

; CURRENT APPLICATION NUMBER: US/09/290,049

; CURRENT FILING DATE: 1999-04-12

; EARLIER APPLICATION NUMBER: 60/081,550

; EARLIER FILING DATE: 1998-04-13

; EARLIER APPLICATION NUMBER: 60/115,142

; EARLIER FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 11

; LENGTH: 21

; TYPE: PRT

; ORGANISM: S. mutans

US-09-290-049-11

Query Match

Best Local Similarity 63.0%; Score 75; DB 16; Length 21;

Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDTPYLHD 21

Db 1 AINHLSILEAMSDNDPOYNKD 21

RESULT 9

US-09-649-885-2

; Sequence 2, Application US/09649885

; GENERAL INFORMATION:

; APPLICANT: Nichols, Scott E.

; TITLE OF INVENTION: Substitutes for Modified Starch and

; FILE REFERENCE: 0358D2

; CURRENT APPLICATION NUMBER: US/09/649,885

; CURRENT FILING DATE: 2000-08-28

; PRIOR APPLICATION NUMBER: US 09/008,172

; PRIOR FILING DATE: 1998-01-16

; PRIOR APPLICATION NUMBER: 08/482,711

; PRIOR FILING DATE: 1995-06-07

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 1430

; TYPE: PRT

; ORGANISM: Streptococcus mutans

US-09-649-885-2

Query Match

Best Local Similarity 71.4%; Score 75; DB 20; Length 1430;

Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANDHLSILEAMSDNDTPYLHD 21

Db 495 AINHLSILEAMSDNDPOYNKD 515

RESULT 10

US-09-740-274-6

; Sequence 6, Application US/09740274

; GENERAL INFORMATION:

; APPLICANT: Nichols, Scott E.

; TITLE OF INVENTION: Glucan-containing Compositions and Paper

; FILE REFERENCE: 0357CRD

; CURRENT APPLICATION NUMBER: US/09/740,274

; CURRENT FILING DATE: 2000-12-19

; PRIOR APPLICATION NUMBER: 09/210,361

; PRIOR FILING DATE: 1998-12-11

```
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
09-740-274-6
```

```
Query Match          63.0%; Score 75; DB 21; Length 1430;
Best Local Similarity 71.4%; Pred. No. 0.034;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 1 ANDHLSLEAMSDNDPTPLHD 21
   1 :|||||
DB 495 AINHLSLEAMSDNDPQYKND 515
```

```
RESULT 11
US-09-290-049-13
; Sequence 13, Application US/09290049
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: FDC98-01p2A
; CURRENT APPLICATION NUMBER: US/09/290,049
; CURRENT FILING DATE: 1999-04-12
; EARLIER APPLICATION NUMBER: 60/081,550
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/115,142
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. downei
US-09-290-049-13
```

```
Query Match          61.3%; Score 73; DB 16; Length 21;
Best Local Similarity 73.7%; Pred. No. 0.00073;
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 ANDHLSLEAMSDNDPTPL 19
   1 :|||||
DB 1 AIDHLSLEAMSGNDNDYV 19
```

```
RESULT 12
US-09-499-203-2
; Sequence 2, Application US/09499203
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNOTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-196P
```

```
; CURRENT APPLICATION NUMBER: US/09/499,203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2057
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-2
```

```
Query Match          50.4%; Score 60; DB 18; Length 2057;
Best Local Similarity 55.0%; Pred. No. 11;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 1 ANDHLSLEAMSDNDPTPLH 20
   1 :|||||
DB 665 ANKHLSDLMWNGKDPQYVN 684
```

```
RESULT 13
US-09-595-298A-661
; Sequence 661, Application US/09595298A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai
; APPLICANT: BROVER, Vyacheslav
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypept
; FILE REFERENCE: 2750-0953P
; CURRENT APPLICATION NUMBER: US/09/595,298A
; CURRENT FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 2756
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 661
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: (1)..(306)
; OTHER INFORMATION: Ceres Seq. ID no. 1026741
; NAME/KEY: misc-feature
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa is any aa, unknown or other
US-09-595-298A-661
```

```
Query Match          42.0%; Score 50; DB 19; Length 306;
Best Local Similarity 52.9%; Pred. No. 49;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 3 DHLSTLEAMSDNDPTPL 19
   1 :|||||
DB 200 DYEQIMEAMSDKGLYV 216
```

```
RESULT 14
US-09-595-298A-660
; Sequence 660, Application US/09595298A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai
; APPLICANT: BROVER, Vyacheslav
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypept
; FILE REFERENCE: 2750-0953P
; CURRENT APPLICATION NUMBER: US/09/595,298A
; CURRENT FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 2756
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 660
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:49:47 ; Search time 87.3 Seconds

(without alignments)  
5.413 Million cell updates/sec

Title: US-09-290-049a-1

Perfect score: 119

Sequence: 1 ANDHSLTLEAMSDNDPRLHD 21

Scoring table: BIOSUM62

Searched: 212252 seqs, 22503292 residues

Number of hits satisfying chosen parameters: 212252

Maximum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents: AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PCRTUS.COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	1475	US-09-007-999-2	Sequence 2, Appl
2	119	100.0	1475	US-09-210-361-2	Sequence 2, Appl
3	110	92.4	1375	US-09-210-361-4	Sequence 4, Appl
4	75	63.0	1430	US-09-008-1172-2	Sequence 2, Appl
5	75	63.0	1430	US-09-210-361-6	Sequence 6, Appl
6	63	52.9	1577	US-08-793-824-2	Sequence 2, Appl
7	45	37.8	385	US-08-361-920-23	Sequence 23, Appl
8	45	37.8	385	US-08-479-939-23	Sequence 23, Appl
9	45	37.8	385	US-08-483-432-23	Sequence 23, Appl
10	44	37.0	384	US-09-311-170-2	Sequence 2, Appl
11	43	36.1	2465	US-08-596-291-3	Sequence 3, Appl
12	43	36.1	2465	US-09-100-804-3	Sequence 3, Appl
13	43	36.1	2465	US-09-080-855-12	Sequence 12, Appl
14	43	36.1	2465	US-09-094-09943-2	Sequence 46, Appl
15	43	36.1	2465	US-09-290-640-46	Sequence 3, Appl
16	42	35.3	259	US-08-277-231A-3	Sequence 3, Appl
17	42	35.3	259	US-08-473-750-6	Sequence 6, Appl
18	42	35.3	259	US-08-477-326-6	Sequence 6, Appl
19	42	35.3	259	US-08-070-301-16	Sequence 16, Appl
20	41	34.5	480	US-08-272-255-12	Sequence 12, Appl
21	41	34.5	480	US-08-933-750C-35	Sequence 35, Appl
22	40	33.6	196	US-08-933-750C-35	Sequence 35, Appl
23	40	33.6	196	US-09-234-613-35	Sequence 35, Appl
24	40	33.6	307	US-08-332-562A-83	Sequence 83, Appl
25	40	33.6	334	US-09-362-473-6	Sequence 6, Appl
26	40	33.6	455	US-09-362-473-6	Sequence 6, Appl
27	40	33.6	761	US-09-625-188-14	Sequence 14, Appl

28	40	33.6	839	4	US-09-197-636-2	Sequence 2, Appl
29	40	33.6	839	4	US-09-197-636-4	Sequence 4, Appl
30	40	33.6	839	4	US-09-197-636-8	Sequence 8, Appl
31	40	33.6	2860	2	US-08-826-267-2	Sequence 2, Appl
32	40	33.6	3898	4	US-08-750-717-2	Sequence 2, Appl
33	39.5	33.2	337	3	US-09-032-372-2	Sequence 2, Appl
34	39.5	33.2	500	6	US-08-258-261B-4	Sequence 4, Appl
35	39.5	33.2	567	1	US-08-456-837-4	Sequence 4, Appl
36	39.5	33.2	567	1	US-08-457-342-4	Sequence 4, Appl
37	39.5	33.2	567	1	US-08-457-646A-4	Sequence 4, Appl
38	39.5	33.2	567	1	US-08-458-076A-4	Sequence 4, Appl
39	39.5	33.2	567	1	US-08-457-335A-4	Sequence 4, Appl
40	39.5	33.2	567	2	US-08-729-214-4	Sequence 4, Appl
41	39.5	33.2	567	2	US-09-028-934-4	Sequence 4, Appl
42	39.5	33.2	567	3	US-09-071-101-2	Sequence 2, Appl
43	39	32.8	755	3	US-09-369-618-2	Sequence 2, Appl
44	39	32.8	755	3	US-09-369-618-2	Sequence 2, Appl
45	39	32.8	755	3	US-09-369-617-2	Sequence 2, Appl

#### ALIGNMENTS

```
RESULT 1
US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0356D
; CURRENT APPLICATION NUMBER: US/09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PR1
; ORGANISM: S. Typhimurium mutans
US-09-007-999-2
Query Match 100.0%; Score 119; DB 3; Length 1475;
Best Local Similarity 100.0%; Pred. No. 6.4e-11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ANDHSLTLEAMSDNDPRLHD 21
DB 481 ANDHSLTLEAMSDNDPRLHD 501
RESULT 2
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
```

```

? EARLIER APPLICATION NUMBER: 09/008,112
? EARLIER FILING DATE: 1998-01-16
? EARLIER APPLICATION NUMBER: 08/482,711
? EARLIER FILING DATE: 1995-06-07
? NUMBER OF SEQ ID NOS: 6
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO: 2
? LENGTH: 1475
? TYPE: prt
? ORGANISM: Streptococcus mutans
? OS: 09-210-361-2

```

Query Match	100.0%	Score 119;	DB 4;	Length 1475;
Best Local Similarity	100.0%	Pred. No. 6.4e-11;		
Matches 21; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

0Y 1 ANDHLSILEAWSNDNDTPYLHD 21  
 |||||  
 481 ANDHLSILEAWSNDNDTPYLHD 501

```

RESULT 3
US-09-210-361-4
Sequence 4, Application US/09210361
Patent No. 6284479
GENERAL INFORMATION:
APPLICANT: Nichols, Scott E.
TITLE OF INVENTION: Substitutes for Modified Starches and
TITLE OF INVENTION: Latexes in Paper Manufacture
FILE REFERENCE: 03570R
CURRENT APPLICATION NUMBER: US/09/210,361
CURRENT FILING DATE: 1998-12-11
EARLIER APPLICATION NUMBER: 09/007,999
EARLIER FILING DATE: 1998-01-16
EARLIER APPLICATION NUMBER: 08/478,704
EARLIER FILING DATE: 1995-06-07
EARLIER APPLICATION NUMBER: 09/009,620
EARLIER FILING DATE: 1998-01-20
EARLIER APPLICATION NUMBER: 08/485,243
EARLIER FILING DATE: 1995-06-07
EARLIER APPLICATION NUMBER: 09/008,172
EARLIER FILING DATE: 1998-01-16
EARLIER APPLICATION NUMBER: 08/482,711
EARLIER FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
30 ID NO 4
LENGTH: 1375
TYPE: PRF
ORGANISM: streptococcus mutans
US-09-210-361-4

```

	Query Match	92.4%	Score 110;	DB 4;	Length 1275;
	Best Local Similarity	95.2%;	Pred. No. 1,7e+09;		
Matches	20; Conservative	0;	Mismatches	1;	Indels 0;
Oy	1 ANDHLSTLEAWSNDPTPLHLD	21			
Db	507 ANDHLSTLEAWSYNDPTPLHD	527			

RESULT 4  
US-09-008-172-2  
Sequence 2, Application US/09008172  
Patent No. 6127602  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Substitutes for Modified Starch and  
TITLE OF INVENTION: latexes in Paper Manufacture  
FILE REFERENCE: 0358D  
CURRENT APPLICATION NUMBER: US/09/008,172

```

:
: CURRENT FILING DATE: 1998-01-16
: EARLIER APPLICATION NUMBER: 08/482,711
: EARLIER FILING DATE: 1995-06-07
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: FASTSEQ For Windows Version 3.0
:
: SEQ ID NO 2
:
: LENGTH: 1430
:
: TYPE: PR1
: ORGANISM: Streptococcus mutans
:
: US-09-008-172-2

```

Query Match	63.0%;	Score 75;	DB 3;	Length 1430;
Best Local Similarity	71.4%;	Pred. No. 0.00084;		
Matches 15;	Conservative 1;	Mismatches 5;	Indels 0;	Gaps 0;

```
QY      1 ANDHLSILEAWSNDNTPYIHD 21
         ↑ : ||||| ||||| ↑ |
Db      495 AINHLSTLEAWSNDNDPQYNKD 515
```

```

RESULT 5
US-09-210-361-6
: Sequence 6, Application US/09210361
: Patent No. 6284479
: GENERAL INFORMATION:
: APPLICANT: Nichols, Scott E.
: TITLE OF INVENTION: Substitutes for Modified Starches and
: TITLE OF INVENTION: Latexes in Paper Manufacture
: FILE REFERENCE: 0357CR
: CURRENT APPLICATION NUMBER: US/09/210,361
: CURRENT FILING DATE: 1998-12-11
: EARLIER APPLICATION NUMBER: 09/007,999
: EARLIER FILING DATE: 1998-01-16
: EARLIER APPLICATION NUMBER: 08/478,704
: EARLIER FILING DATE: 1995-06-07
: EARLIER APPLICATION NUMBER: 09/009,620
: EARLIER FILING DATE: 1998-01-20
: EARLIER APPLICATION NUMBER: 08/485,243
: EARLIER FILING DATE: 1995-06-07
: EARLIER APPLICATION NUMBER: 09/008,172
: EARLIER FILING DATE: 1998-01-16
: EARLIER APPLICATION NUMBER: 08/482,711
: EARLIER FILING DATE: 1995-06-07
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 6
: LENGTH: 1430
: TYPE: PRT
: ORGANISM: Streptococcus mutans
US-09-210-361-6

```

Query Match	63.0%	Score 75	DB 4	Length 1430
Best Local Similarity	71.4%	Pred. No.	0.00084	
Matches	15	Conservative	1	Mismatches 5; Indels 0; Gaps 0
Qy	1	ANDHSLLEAASDNDPTFLHD	21	
		: : : : :   : : : : :		
Db	495	AINHSLLEAASDNDPQYRKD	515	

RESULT 6  
 US-08-793-824-2  
 : Sequence 2, Application US/08793824  
 : Patent No. 5981838  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: Simpson, Christine Lynn  
 : APPLICANT: Giffard, Philip Morrison  
 : APPLICANT: Jacques, Nicholas Anthony  
 : TITLE OF INVENTION: Genetic Manipulation of Plants to  
 : TITLE OF INVENTION: Increase Stored Carbohydrates  
 : NUMBER OF SEQUENCES: 2









```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/09943
: FILING DATE: 01-SEP-1994
: CLASSIFICATION:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/115,573
: FILING DATE: 01-SEP-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: TWOMEY, MICHAEL J.
: REGISTRATION NUMBER: P-38,349
: REFERENCE/DOCKET NUMBER: L0461/7000WO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/720-3500
: TELEFAX: 617/720-2441
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2466 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US94-09943-2

```

```

Query Match          36.1%; Score 43; DB 5; Length 2466;
Best Local Similarity 56.2%; Pred. No. 2.5e+02;
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

QY 4 HLSILE--AMSDNDTP 17
   ||| |||:|
Db 2348 HISHLNTAWPDHDT 2363

```

```

RESULT 15
US-09-290-640-46
: Sequence 46, Application US/09290640
: Patent No. 6204055
: GENERAL INFORMATION:
: APPLICANT: Dean, Nicholas M.
: APPLICANT: Marcussen, Eric G.
: TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signaling
: FILE REFERENCE: ISPH-0351
: CURRENT APPLICATION NUMBER: US/09/290,640
: CURRENT FILING DATE: 1999-04-12
: NUMBER OF SEQ ID NOS: 85
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 46
: LENGTH: 2485
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-290-640-46

```

```

Query Match          36.1%; Score 43; DB 4; Length 2485;
Best Local Similarity 56.2%; Pred. No. 2.5e+02;
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

QY 4 HLSILE--AMSDNDTP 17
   ||| |||:|
Db 2367 HISHLNTAWPDHDT 2382

```

Search completed: March 27, 2002, 13:59:26  
Job time: 579 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:54:33 ; Search time 137.48 Seconds  
(without alignments)  
10.540 Million cell updates/sec

Title: US-09-290-049A-1  
Perfect score: 119  
Sequence: 1 ANDHLSILEAWSNDPTPLYLD 21

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 332938 seqs, 6899538 residues

Number of hits satisfying chosen parameters: 332938

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_New:\*  
1: /cgn2-6/ptodata/1/paa/PCr\_NEW\_COMB.pep:\*  
2: /cgn2-6/ptodata/1/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2-6/ptodata/1/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2-6/ptodata/1/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2-6/ptodata/1/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2-6/ptodata/1/paa/US10\_NEW\_COMB.pep1:\*  
7: /cgn2-6/ptodata/1/paa/US10\_NEW\_COMB.pep:\*  
8: /cgn2-6/ptodata/1/paa/US60\_NEW\_COMB.pep:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	21	US-09-562-328-20	Sequence 20, Appl
2	119	100.0	21	US-09-290-049A-1	Sequence 1, Appl
3	119	100.0	21	US-09-290-049A-10	Sequence 10, Appl
4	110	92.4	21	US-09-562-328-21	Sequence 21, Appl
5	108	90.8	21	US-09-562-328-23	Sequence 23, Appl
6	108	90.8	21	US-09-562-328-25	Sequence 25, Appl
7	108	90.8	21	US-09-290-049A-12	Sequence 12, Appl
8	108	90.8	21	US-09-290-049A-14	Sequence 14, Appl
9	75	63.0	21	US-09-562-328-22	Sequence 22, Appl
10	75	63.0	21	US-09-290-049A-11	Sequence 11, Appl
11	75	63.0	21	US-09-604-957-4	Sequence 4, Appl
12	73	61.3	21	US-09-562-328-24	Sequence 24, Appl
13	73	61.3	21	US-09-290-049A-13	Sequence 13, Appl
14	71	59.7	523	US-09-604-957-5	Sequence 5, Appl
15	60	50.4	584	US-09-604-957-6	Sequence 6, Appl
16	49.5	41.6	423	US-09-708-427-55326	Sequence 55326, A
17	49.5	41.6	511	US-09-708-427-55325	Sequence 55325, A
18	49.5	41.6	549	US-09-708-427-55324	Sequence 55324, A
19	49	41.2	302	PCT-US02-03987-14004	Sequence 14004, A
20	49	41.2	302	US-09-815-242-14004	Sequence 14004, A
21	49	41.2	302	US-10-072-851-14004	Sequence 14004, A
22	49	41.2	535	US-09-604-957-7	Sequence 7, Appl
23	49	41.2	1278	US-09-604-957-3	Sequence 3, Appl
24	47.5	39.9	434	US-09-708-427-7994	Sequence 7994, Ap
25	47.5	39.9	519	US-09-708-427-7993	Sequence 7993, Ap

26	47	39.5	15	7	US-10-067-649-90	Sequence 90, Appl
27	47	39.5	366	6	US-09-708-427-27415	Sequence 27415, A
28	47	39.5	367	6	US-09-708-427-27415	Sequence 27415, A
29	47	39.5	378	6	US-09-708-427-27414	Sequence 27414, A
30	46.5	39.1	1604	6	US-09-888-615-73	Sequence 73, Appl
31	46	38.7	199	6	US-09-605-703B-2226	Sequence 2226, Ap
32	46	38.7	2590	6	US-09-897-516-7203	Sequence 7203, Ap
33	46	38.7	2603	6	US-09-897-516-5891	Sequence 5891, Ap
34	45	37.8	185	6	US-09-897-516-5719	Sequence 5719, Ap
35	45	37.8	236	6	US-09-154-750A-80	Sequence 80, Appl
36	45	37.8	638	6	US-09-708-427-25322	Sequence 25322, A
37	45	37.8	765	6	US-09-708-427-25321	Sequence 25321, A
38	45	37.8	947	6	US-09-708-427-25320	Sequence 25320, A
39	44.5	37.4	499	6	US-09-614-150-42525	Sequence 42525, A
40	44	37.0	322	1	PCT-US02-03987-11634	Sequence 11634, A
41	44	37.0	322	1	US-09-815-242-11634	Sequence 11634, A
42	44	37.0	322	7	US-10-072-851-11634	Sequence 11634, A
43	44	37.0	331	1	PCT-US02-03987-10188	Sequence 10188, A
44	44	37.0	331	6	US-09-815-242-10188	Sequence 10188, A
45	44	37.0	331	7	US-10-072-851-10188	Sequence 10188, A

#### ALIGNMENTS

RESULT 1  
US-09-562-328-20  
Sequence 20, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 0495, 0046-01  
CURRENT FILING DATE: US/09/562,328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 20  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-20

Query Match 100.0%; Score 119; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2,9e-12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ANDHLSILEAWSNDPTPLYLD 21  
DB 1 ANDHLSILEAWSNDPTPLYLD 21  
RESULT 2  
US-09-290-049A-1  
Sequence 1, Application US/09290049A  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
FILE REFERENCE: 1564, 1008-002  
CURRENT APPLICATION NUMBER: US/09/290,049A  
CURRENT FILING DATE: 1999-04-12  
PRIOR APPLICATION NUMBER: 60/081,550  
PRIOR FILING DATE: 1998-04-13  
PRIOR APPLICATION NUMBER: 60/115,142  
PRIOR FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: EAW peptide  
US-09-290-049a-1

Query Match 100.0%; Score 119; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2,9e-12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDPTPYLHD 21  
|||  
DB 1 ANDHLSILEAWSNDPTPYLHD 21

ULR 3  
09-290-049a-10  
Sequence 10, Application US/09290049A  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.  
APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
FILE REFERENCE: 1564.1008-002  
CURRENT APPLICATION NUMBER: US/09/290,049A  
CURRENT FILING DATE: 1999-04-12  
PRIOR APPLICATION NUMBER: 60/081,550  
PRIOR FILING DATE: 1998-04-13  
PRIOR APPLICATION NUMBER: 60/115,142  
PRIOR FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10  
LENGTH: 21  
TYPE: PRT  
ORGANISM: S. mutans  
US-09-290-049a-10

Query Match 100.0%; Score 119; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2,9e-12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ANDHLSILEAWSNDPTPYLHD 21  
|||  
1 ANDHLSILEAWSNDPTPYLHD 21

RESULT 4  
US-09-562-328-21

Sequence 21, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
APPLICANT: SMITH, DANIEL J.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARRIES  
FILE REFERENCE: 04995.0046-01  
CURRENT APPLICATION NUMBER: US/09/562,328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 21  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-21

Query Match 92.4%; Score 110; DB 6; Length 21;  
Best Local Similarity 95.2%; Pred. No. 7,4e-11;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDPTPYLHD 21  
|||  
DB 1 ANDHLSILEAWSNDPTPYLHD 21

RESULT 5  
US-09-562-328-23  
Sequence 23, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
APPLICANT: SMITH, DANIEL J.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARRIES  
FILE REFERENCE: 04995.0046-01  
CURRENT APPLICATION NUMBER: US/09/562,328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 23  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-23

Query Match 90.8%; Score 108; DB 6; Length 21;  
Best Local Similarity 85.7%; Pred. No. 1,5e-10;  
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDPTPYLHD 21  
|||  
DB 1 ANNHVSIVEAWSNDPTPYLHD 21

RESULT 6  
US-09-562-328-25  
Sequence 25, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.  
APPLICANT: SMITH, DANIEL J.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARRIES  
FILE REFERENCE: 04995.0046-01  
CURRENT APPLICATION NUMBER: US/09/562,328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 25  
LENGTH: 21  
TYPE: PRT  
ORGANISM: Streptococcus sp.  
US-09-562-328-25

Query Match 90.8%; Score 108; DB 6; Length 21;  
Best Local Similarity 85.7%; Pred. No. 1,5e-10;  
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDPTPYLHD 21  
|||  
DB 1 ANNHVSIVEAWSNDPTPYLHD 21

RESULT 7  
US-09-290-049a-12

```
; Sequence 12, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290.049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. downei
US-09-290-049A-12
```

```
Query Match          90.8%; Score 108; DB 6; Length 21;
Best Local Similarity 85.7%; Pred. No. 1.5e-10;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 ANDHLSILEAMSDNDPTPLHD 21
DB      1 AINHVSIYEAMSDNDPTPLHD 21
```

```
RESULT      8
US-09-290-049A-14
; Sequence 14, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290.049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. sobrinus
US-09-290-049A-14
```

```
Query Match          90.8%; Score 108; DB 6; Length 21;
Best Local Similarity 85.7%; Pred. No. 1.5e-10;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 ANDHLSILEAMSDNDPTPLHD 21
DB      1 AINHVSIYEAMSDNDPTPLHD 21
```

```
RESULT      9
US-09-562-328-22
; Sequence 22, Application US/09562328
; GENERAL INFORMATION:
; APPLICANT: LEES, ANDREW
; APPLICANT: TAUBMAN, MARTIN A.
; APPLICANT: SMITH, DANIEL J.
; TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES
; FILE REFERENCE: 04995.0046-01
```

```
; CURRENT APPLICATION NUMBER: US/09/562.328
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 09/288,965
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Streptococcus sp.
US-09-562-328-22
```

```
Query Match          63.0%; Score 75; DB 6; Length 21;
Best Local Similarity 71.4%; Pred. No. 2.3e-05;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
```

```
OY      1 ANDHLSILEAMSDNDPTPLHD 21
DB      1 AINHLSTLEAMSDNDPQYNKD 21
```

```
RESULT      10
US-09-290-049A-11
; Sequence 11, Application US/09290049A
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290.049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. mutans
US-09-290-049A-11
```

```
Query Match          63.0%; Score 75; DB 6; Length 21;
Best Local Similarity 71.4%; Pred. No. 2.3e-05;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
```

```
OY      1 ANDHLSILEAMSDNDPTPLHD 21
DB      1 AINHLSTLEAMSDNDPQYNKD 21
```

```
RESULT      11
US-09-604-957-4
; Sequence 4, Application US/09604957
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUTI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604.957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 545
```

TYPE: PRF  
ORGANISM: Streptococcus mutans  
US-09-604-957-4

Query Match 63.0%; Score 75; DB 6; Length 545;  
Best Local Similarity 71.4%; Pred. No. 0.00094;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANDHLSLEAMSDNDPTVYL 21  
DB 75 ANHLSLEAMSDNDPQYND 95

RESULT 12  
US-09-562-328-24  
Sequence 24, Application US/09562328  
GENERAL INFORMATION:  
APPLICANT: LEES, ANDREW  
APPLICANT: TAUBMAN, MARTIN A.

APPLICANT: SMITH, DANIEL J.  
TITLE OF INVENTION: CONJUGATE VACCINES FOR THE PREVENTION OF DENTAL CARIES  
FILE REFERENCE: 04995.0046-01  
CURRENT APPLICATION NUMBER: US/09/562,328  
CURRENT FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 09/288,965  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 24  
LENGTH: 21  
TYPE: PRF  
ORGANISM: Streptococcus sp.  
US-09-562-328-24

Query Match 61.3%; Score 73; DB 6; Length 21;  
Best Local Similarity 73.7%; Pred. No. 4.7e-05;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANDHLSLEAMSDNDPTVYL 19  
DB 1 AIDHLSLEAMSGNDNDYV 19

RESULT 13  
US-09-290-049A-13  
Sequence 13, Application US/09290049A  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel J.

APPLICANT: Taubman, Martin A.  
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
FILE REFERENCE: 1564.1008-002  
CURRENT APPLICATION NUMBER: US/09/290,049A  
CURRENT FILING DATE: 1999-04-12  
PRIOR APPLICATION NUMBER: 60/081,550  
PRIOR FILING DATE: 1998-04-13  
PRIOR APPLICATION NUMBER: 60/115,142  
PRIOR FILING DATE: 1999-01-08  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13  
LENGTH: 21  
TYPE: PRF  
ORGANISM: S. downei  
US-09-290-049A-13

Query Match 61.3%; Score 73; DB 6; Length 21;  
Best Local Similarity 73.7%; Pred. No. 4.7e-05;  
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 ANDHLSLEAMSDNDPTVYL 19  
DB 1 AIDHLSLEAMSGNDNDYV 19

RESULT 14  
US-09-604-957-5  
Sequence 5, Application US/09604957  
GENERAL INFORMATION:  
APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
APPLICANT: DIJKHUIZEN, LOBBERT  
APPLICANT: RAHAOUI, HAKIM  
APPLICANT: LEER, ROBERT-JAN  
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
FILE REFERENCE: BO 43388  
CURRENT APPLICATION NUMBER: US/09/604,957  
CURRENT FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 523  
TYPE: PRF  
ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-5

Query Match 59.7%; Score 71; DB 6; Length 523;  
Best Local Similarity 66.7%; Pred. No. 0.0038;  
Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANDHLSLEAMSDNDPTVYL 21  
DB 75 ANHLSLEAMSHNDPLVYTD 95

RESULT 15  
US-09-604-957-6  
Sequence 6, Application US/09604957  
GENERAL INFORMATION:  
APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
APPLICANT: DIJKHUIZEN, LOBBERT  
APPLICANT: RAHAOUI, HAKIM  
APPLICANT: LEER, ROBERT-JAN  
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
FILE REFERENCE: BO 43388  
CURRENT APPLICATION NUMBER: US/09/604,957  
CURRENT FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 00201871.1  
PRIOR FILING DATE: 2000-05-25  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 6  
LENGTH: 584  
TYPE: PRF  
ORGANISM: Leuconostoc mesenteroides  
US-09-604-957-6

Query Match 50.4%; Score 60; DB 6; Length 584;  
Best Local Similarity 55.0%; Pred. No. 0.23;  
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANDHLSLEAMSDNDPTVYL 20  
DB 75 ANHLSLEAMSGNDPQYVN 94

Search completed: March 27, 2002, 14:22:45  
Job time: 1692 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:53:18 ; Search time 102.51 Seconds  
(without alignments)  
15.605 Million cell updates/sec

Title: US-09-290-049a-1

Sequence: 1 ANDHLSILEWMSDNDPTPLHD 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	100.0	1475	2 B33135	gtfB protein precu
2	110	92.4	1375	2 JT0345	dextranucrase (EC
3	108	90.8	1592	2 A38175	glucosyltransferas
4	78	65.5	1365	2 A41483	glucosyltransferas
5	75	63.0	1431	2 A45866	dextranucrase (EC
6	74	62.2	1508	2 T31098	probable dextranu
7	66	55.5	1449	2 T30857	glucosyltransferas
8	66	55.5	1449	2 T30552	glucosyltransferas
9	63	52.9	1577	2 T30858	glucosyltransferas
10	58	48.7	1518	2 A44811	glucosyltransferas
11	57	47.9	331	2 B48445	glyceraledehyde-3-p
12	57	47.9	1599	2 S22737	glucosyltransferas
13	52	43.7	4848	2 T30289	pristinamycin I sy
14	50	42.0	347	2 T48610	hypothetical prote
15	49.5	41.6	349	2 T03963	rf2 nuclear restor
16	49	41.2	336	1 Q0BE40	BGLF2 protein - hu
17	48	40.3	490	2 H70538	probable ppdk prot
18	47.5	39.9	331	1 DEUTC6	glyceraledehyde-3-p
19	47.5	39.9	519	2 C86372	hypothetical prote
20	47	39.5	175	2 C86205	hypothetical prote
21	47	39.5	378	2 T04254	hypothetical prote
22	47	39.5	524	2 D82220	conserved hypotet
23	46	38.7	275	2 T40088	RhogeF domain cont
24	46	38.7	275	2 S55978	hypothetical prote
25	46	38.7	335	2 E86568	glyceraledehyde-3-p
26	46	38.7	335	2 B22053	glyceraledehyde 3-p
27	46	38.7	476	1 S58229	salicylate biosynt
28	45.5	38.2	96	2 E81786	conserved hypotet
29	45	37.8	236	2 S62732	guanidinacetate N

30	45	37.8	947	2 E86362	hypothetical prote
31	45	37.8	1159	2 T43461	probable phosphodi
32	44.5	37.4	210	2 H83332	conserved hypotet
33	44	37.0	247	2 P00178	glyceraledehyde-3-p
34	44	37.0	331	1 DEECG3	glyceraledehyde-3-p
35	44	37.0	331	2 D85788	glyceraledehyde-3-p
36	44	37.0	333	1 DECHG3	glyceraledehyde-3-p
37	44	37.0	333	2 JC5370	glyceraledehyde-3-p
38	44	37.0	337	1 DEJUC6	glyceraledehyde-3-p
39	44	37.0	337	1 DEZMGC	glyceraledehyde-3-p
40	44	37.0	337	1 DEBHG	glyceraledehyde-3-p
41	44	37.0	436	2 T24953	hypothetical prote
42	44	37.0	468	2 S61964	probable membrane
43	44	37.0	1208	2 B82091	exodexyribonuclea
44	44	37.0	1313	2 T29027	hypothetical prote
45	43.5	36.6	97	2 S12958	tachykinin delta p

## ALIGNMENTS

```

RESULT 1
B33135
gtfB protein precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 15-Oct-1999
C:Accession: B33135; A33128
R:Shiroza, T.; Ueda, S.; Kuramitsu, H. K.
J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A:Reference number: A33135; MUID:87308013
A:Accession: B33135
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1475 <SH1>
A:Cross-references: GB:M17361; NID:g153639; PID:AAA8588.1; PID:g153640
R:Shiroza, T.; Ueda, S.; Kuramitsu, H. K.
submitted to the Protein Sequence Database, September 1990
A:Reference number: A33128
A:Accession: A33128
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-171, 173-641, 'N', 643-1475 <SH2>
A:Experimental source: strain GS-5
C:Superfamily: cpl repeat homology
F:1096-1115/Domain: cpl repeat homology <CP1>
F:1124-1243/Domain: cpl repeat homology <CP2>
F:1289-1308/Domain: cpl repeat homology <CP3>
F:1354-1373/Domain: cpl repeat homology <CP4>
F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match          100.0%; Score 119; DB 2; Length 1475;
Best Local Similarity 100.0%; Pred. No. 3.9e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0;
OY 1 ANDHLSILEWMSDNDPTPLHD 21
DB 481 ANDHLSILEWMSDNDPTPLHD 501

RESULT 2
JT0345
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)
N:Alternate names: sucrose 6-glucosyltransferase
C:Species: Streptococcus mutans
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 24-Sep-1999
C:Accession: JT0345; C33135
R:Ueda, S.; Shiroza, T.; Kuramitsu, H. K.
Gene 69, 101-109, 1988
A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.
A:Reference number: JT0345; MUID:89137980
A:Accession: JT0345

```

A:Molecule type: DNA  
A:Residues: 1-1375 <UED>  
C:Accession: A41483  
R:Gillmore, K.S.; Russell, R.R.B.; Ferretti, J.J.  
J. Infect. Immun. 58, 2452-2458, 1990  
A:Title: Analysis of the Streptococcus downei gtfB gene, which specifies a glucosyltransferase  
A:Reference number: A41483; MUID:90316665  
A:Accession: A41483  
A:Molecule type: DNA  
A:Residues: 1-349 <SH1>  
A:Cross-References: GB:M17361  
C:Genetics:  
A:Gene: gtfC  
C:Function:  
A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans  
C:Superfamily: cpl repeat homology  
C:Keywords: duplication; glucosyltransferase; hexosyltransferase  
-34/Domain: signal sequence #status predicted <SIG>  
-35-1375/Product: glucosyltransferase #status predicted <MAT>  
F:1126-1145/Domain: cpl repeat homology <CP1>  
F:1253-1272/Domain: cpl repeat homology <CP2>  
F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 92.4%; Score 110; DB 2; Length 1375;  
Best Local Similarity 95.2%; Pred. No. 9, 1e-09;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDTPYLHD 21  
|||||  
Db 507 ANDHLSILEAWSNDTPYLHD 527

RESULT 3  
A38175  
glucosyltransferase precursor - Streptococcus sobrinus  
C:Species: Streptococcus sobrinus  
C>Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 15-Oct-1999  
C:Accession: A38175  
R:Abdo, H.; Msumura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.  
J. Bacteriol. 173, 989-996, 1991  
A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus  
A:Reference number: A38175; MUID:91123227  
A:Accession: A38175  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1592 <ABO>  
A:Cross-References: GB:D90213; NID:9217032; PIDN:BAAL4241.1; PID:01014946; PID:9217033  
C:Superfamily: cpl repeat homology  
F:1093-1112/Domain: cpl repeat homology <CP1>  
F:1222-1241/Domain: cpl repeat homology <CP2>  
F:1287-1306/Domain: cpl repeat homology <CP3>  
F:1330-1351/Domain: cpl repeat homology <CP4>  
F:1352-1371/Domain: cpl repeat homology <CP5>  
F:1402-1420/Domain: cpl repeat homology <CP6>  
F:1465-1484/Domain: cpl repeat homology <CP7>  
F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 90.8%; Score 108; DB 2; Length 1592;  
Best Local Similarity 85.7%; Pred. No. 2, 2e-08;  
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDTPYLHD 21  
|||||  
Db 477 ANNHVSIVEAWSNDTPYLHD 497

RESULT 4  
A41483  
glucosyltransferase (EC 2.4.1.-) gtfS precursor - Streptococcus sobrinus  
C:Species: Streptococcus sobrinus

C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999  
C:Accession: A41483  
R:Gillmore, K.S.; Russell, R.R.B.; Ferretti, J.J.  
Infect. Immun. 58, 2452-2458, 1990  
A:Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltransferase  
A:Reference number: A41483; MUID:90316665  
A:Accession: A41483  
A:Molecule type: DNA  
A:Residues: 1-1365 <GIL>  
A:Cross-References: GB:M30943; NID:9153652; PIDN:AAA26898.1; PID:9153653  
C:Genetics:  
A:Gene: gtfS  
C:Superfamily: cpl repeat homology  
C:Keywords: glucosyltransferase; hexosyltransferase

Query Match 65.5%; Score 78; DB 2; Length 1365;  
Best Local Similarity 71.4%; Pred. No. 0.00087;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDTPYLHD 21  
|||||  
Db 467 AIDHLSILEAWSNDTPYLHD 487

RESULT 5  
A45866  
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans  
C:Species: Streptococcus mutans  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: A45866  
R:Honda, O.; Kato, C.; Kuramitsu, H.K.  
J. Gen. Microbiol. 136, 2099-2105, 1990  
A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the gluco  
A:Reference number: A45866; MUID:91100958  
A:Accession: A45866  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1431 <HON>  
A:Cross-References: GB:M29296  
C:Superfamily: cpl repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase  
F:181-201/Domain: cpl repeat homology <CP1>  
F:1127-1146/Domain: cpl repeat homology <CP2>  
F:1192-1211/Domain: cpl repeat homology <CP3>  
F:1257-1276/Domain: cpl repeat homology <CP4>  
F:1277-1297/Domain: cpl repeat homology <CP5>  
F:1321-1340/Domain: cpl repeat homology <CP6>  
F:1341-1361/Domain: cpl repeat homology <CP7>  
F:1385-1404/Domain: cpl repeat homology <CP8>

Query Match 63.0%; Score 75; DB 2; Length 1431;  
Best Local Similarity 71.4%; Pred. No. 0.0027;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 ANDHLSILEAWSNDTPYLHD 21  
|||||  
Db 495 AINHLSILEAWSNDTPYLHD 515

RESULT 6  
T31098  
probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides  
C:Species: Leuconostoc mesenteroides  
C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
C:Accession: T31098  
R:Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willemot, R.M.  
FEMS Microbiol. Lett. 159, 307-315, 1998  
A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase  
A:Reference number: Z20981; MUID:96164374  
A:Accession: T31098  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
A:Residues: 1-1508 <MON>  
A:Cross-references: EMBL:AF030129; NID:g2766611; PID:g2766612; PIDN:AAB95453.1  
A:Experimental source: strain NRRL B-1299  
C:Genetics:  
A:Gene: dsrB  
C:Function:  
A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 62.2%; Score 74; DB 2; Length 1508;  
Best Local Similarity 66.7%; Pred. No. 0.0041;  
Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 ANDHSLTEAMSDNDTPYLHD 21  
||| ||||| ||| ||| |||  
Db 563 ANOHLSTLEDMWSHNDPEYKXD 583

RESULT 7  
glycosyltransferase - Streptococcus salivarius

C:Species: Streptococcus salivarius  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30857  
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
Infect. Immun. 63, 609-621, 1995  
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pr  
A:Reference number: 220909; MUID:95122197  
A:Accession: T30857  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1449 <SIM>  
A:Cross-references: EMBL:L35495; NID:g662378; PID:g662379; PIDN:AAC41412.1  
C:Genetics:  
A:Gene: gltI

Query Match 55.5%; Score 66; DB 2; Length 1449;  
Best Local Similarity 72.2%; Pred. No. 0.069;  
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 HSLTEAMSDNDTPYLHD 21  
||| ||||| ||| ||| |||  
Db 539 HSLTEAMSHNDAYNED 556

RESULT 8  
glycosyltransferase N - Streptococcus salivarius (fragment)

C:Species: Streptococcus salivarius  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30552  
R:Jaaffe, R.I.  
submitted to the EMBL Data Library, February 1998  
A:Description: Streptococcus salivarius VI477 gltN.  
A:Reference number: 220854  
A:Accession: T30552  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1449 <JAF>  
A:Cross-references: EMBL:AF049609; NID:g2935545; PID:g2935546; PIDN:AAC05156.1  
C:Genetics:  
A:Gene: gltN

Query Match 55.5%; Score 66; DB 2; Length 1449;  
Best Local Similarity 72.2%; Pred. No. 0.069;  
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 HSLTEAMSDNDTPYLHD 21  
||| ||||| ||| ||| |||

Db 539 HSLTEAMSHNDAYNED 556

RESULT 9  
glycosyltransferase - Streptococcus salivarius

C:Species: Streptococcus salivarius  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30858  
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
Infect. Immun. 63, 609-621, 1995  
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for  
A:Reference number: 220909; MUID:95122197  
A:Accession: T30858  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1577 <SIM>  
A:Cross-references: EMBL:L35928; NID:g662380; PID:g662381; PIDN:AAC41413.1  
C:Genetics:  
A:Gene: gltM

Query Match 52.9%; Score 63; DB 2; Length 1577;  
Best Local Similarity 72.2%; Pred. No. 0.23;  
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 HSLTEAMSDNDTPYLHD 21  
||| ||||| ||| ||| |||  
Db 591 HSLTEAMSYNDHOYKND 608

RESULT 10

glycosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius  
C:Species: Streptococcus salivarius  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 15-Oct-1999  
C:Accession: A44811; S22726; S28809  
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
J. Gen. Microbiol. 137, 2577-2593, 1991  
A:Title: Molecular characterization of a cluster of at least two glycosyltransferase  
A:Reference number: A44811; MUID:92148377  
A:Accession: A44811  
A:Molecule type: DNA  
A:Residues: 1-1518 <GIF>  
A:Cross-references: EMBL:211873; NID:g47526; PIDN:CAA77900.1; PID:g47527  
A:Note: sequence extracted from NCBI backbone (NCBIN:81050, NCBI:81052)  
C:Genetics:  
A:Gene: gltJ  
C:Superfamily: (2) repeat homology  
C:Keywords: glycosyltransferase; hexosyltransferase  
F11307-1326/DomTm: cpi repeat homology <CP4>

Query Match 48.7%; Score 58; DB 2; Length 1518;  
Best Local Similarity 66.7%; Pred. No. 1.3;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 4 HSLTEAMSDNDTPY 18  
||| ||||| ||| ||| |||  
Db 504 HSLTEAMSLNDNHX 518

RESULT 11

glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12) - Leishmania mexicana  
C:Species: Leishmania mexicana  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 11-Jun-1999  
C:Accession: B48445; S25142  
R:Hannaert, V.; Blaauw, M.; Kohl, E.; Allert, S.; Opperdoes, F.R.; Michels, P.A.M.  
Mol. Biochem. Parasitol. 55, 115-126, 1992  
A:Title: Molecular analysis of the cytosolic and glycosomal glyceraldehyde-3-phosphat  
A:Reference number: A48445; MUID:93063042  
A:Accession: B48445

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-331 <HNA>  
A:Cross-references: EMBL:X65220; NID:g9552; PIDD:CAA46323.1; PID:g9553  
C:Superfamily: glyceralddehyde-3-phosphate dehydrogenase  
C:Keywords: oxidoreductase

Query Match 47.9%; Score 57; DB 2; Length 331;  
Best Local Similarity 52.6%; Pred. No. 0.31;  
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 NDHLSLEAWSNDNTPYLH 20  
||| | : | | | | |  
Db 301 NDHFWKLVSWYDNETGSH 319

RESULT 12

2737

glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius

C:Species: Streptococcus salivarius

C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000

C:Accession: S22737; S28810; B44811; S22727

R:Jacques, N.

submitted to the EMBL Data Library, March 1992

A:Reference number: S22726

A:Accession: S22737

A:Molecule type: DNA

A:Residues: 1-1599 <JMC>

A:Cross-references: EMBL:Z11872; NID:g47530; PIDD:CAA7898.1; PID:g47531

A:Experimental source: ATCC 25975

R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.

J. Gen. Microbiol. 137, 2577-2593, 1991

A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes

A:Reference number: A44811; MUID:92148377

A:Accession: S28810

A:Molecule type: DNA

A:Residues: 1-51 <GIF>

A:Cross-references: EMBL:Z11873

C:Genetics:

A:Gene: gtfK

C:Superfamily: cpl repeat homology

C:Keywords: glycosyltransferase; hexosyltransferase

F:1456-1475/Domain: cpl repeat homology <CPR>

Query Match 47.9%; Score 57; DB 2; Length 1599;

Best Local Similarity 55.6%; Pred. No. 2;  
Matches 10; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 4 HLSLEAWSNDNTPYLH 21  
:::||||| | | | |  
Db 494 NISILEAWSNDPYVNE 511

RESULT 13

730289

pristinamycin I synthase 3 - Streptomyces pristinaespiralis

C:Species: Streptomyces pristinaespiralis

C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 01-Dec-2000

C:Accession: T30289

R:de Grey-Lagard, V.A.; Saurin, W.; Thibaut, D.; Gill, P.; Naudin, L.; Crouzet, J.; Bian

submitted to the EMBL Data Library, February 1997

A:Description: Streptogramin B biosynthesis in Streptomyces pristinaespiralis and Strept

A:Reference number: Z20808

A:Accession: T30289

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-4848 <DEC>

A:Cross-references: EMBL:Y11548; NID:e1025755; PID:e307539; PIDD:CAA72312.1

C:Genetics:

A:Gene: snbDE

C:Superfamily: acyl carrier protein homology; acetate--CoA ligase homology

C:Keywords: carrier protein; phosphopantetheine; phosphoprotein

F:511-951/Domain: acetate--CoA ligase homology <ACL1>

F:966-1034/Domain: acyl carrier protein homology <ACP1>

F:1563-2024/Domain: acetate--CoA ligase homology #status atypical <ACL2>

F:2398-2467/Domain: acyl carrier protein homology <ACP2>

F:2995-3427/Domain: acetate--CoA ligase homology <ACP3>

F:3441-3509/Domain: acyl carrier protein homology <ACP3>

F:4043-4492/Domain: acetate--CoA ligase homology <ACL4>

F:4507-4575/Domain: acyl carrier protein homology <ACP4>

F:998,2431,3473/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 43.7%; Score 52; DB 2; Length 4848;  
Best Local Similarity 47.1%; Pred. No. 44;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 ANDHLSLEAWSNDNTP 17  
| : | | : | | | | |  
Db 3972 ADDHARVLTGMWMDTDP 3988

RESULT 14

748610

hypothetical protein F18022.160 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T48610

R:Byran, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.

submitted to the Protein Sequence Database, April 2000

A:Reference number: Z24493

A:Accession: T48610

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-347 <BEV>

A:Cross-references: EMBL:DB

A:Experimental source: cultivar Columbia; BAC clone F18022

C:Genetics:

A:Map position: 5

A:Introns: 232/3; 296/3

A:Note: F18022.160

Query Match 42.0%; Score 50; DB 2; Length 347;  
Best Local Similarity 52.9%; Pred. No. 4;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 DHLSLEAWSNDNTPYL 19  
| : | | | | | | | | |  
Db 200 DYEQIWEAWSKXTLYV 216

RESULT 15

703983

rf2 nuclear restorer protein - maize

N:Alternate names: aldehyde dehydrogenase homolog

C:Species: Zea mays (maize)

C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000

C:Accession: T03983

R:Cul, X.; Wise, R.P.; Schnable, P.S.

Science 272, 1334-1336, 1996

A:Title: The rf2 nuclear restorer gene of male-sterile T-cytoplasm maize.

A:Reference number: Z15177; MUID:96243131

A:Accession: T03983

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-549 <CU1>

A:Cross-references: EMBL:U43082; NID:g1421729; PIDD:AAC49371.1; PID:g1421730

A:Experimental source: strain B73

C:Genetics:

A:Map position: 9

A:Note: rf2

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology  
F:107-370/Domain: aldehyde dehydrogenase homology <ALD>

Query Match 41.6%; Score 49.5; DB 2; Length 549;  
Best Local Similarity 58.8%; Pred. No. 8.2;  
Matches 10; Conservative 1; Mismatches 5; Indels 1; Gaps 1;  
QY 2 NDHLSILEAWSNDNDTPY 18  
|||: ||| |||  
Db 147 NDELALETW-DNGKPY 162

Search completed: March 27, 2002, 14:01:15  
Job time: 477 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:59:26 ; Search time 87.3 Seconds  
(without alignments)  
4.898 Million cell updates/sec

Title: US-09-290-049a-2  
Perfect score: 95  
Sequence: 1 VPSYSFIRAHDSVODLIA 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Number of hits satisfying chosen parameters: 212252

Maximum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCRTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	95	100.0	1475	3 US-09-007-999-2	Sequence 2, Appli
2	95	100.0	1475	4 US-09-210-361-2	Sequence 2, Appli
3	91	95.8	1375	4 US-09-210-361-4	Sequence 4, Appli
4	65	68.4	1430	3 US-09-008-172-2	Sequence 2, Appli
5	65	68.4	1430	4 US-09-210-361-6	Sequence 6, Appli
6	61	64.2	1577	2 US-08-793-824-2	Sequence 2, Appli
7	42	44.2	2627	2 US-08-751-189-3	Sequence 3, Appli
8	42	44.2	2627	2 US-09-060-836-3	Sequence 3, Appli
9	42	44.2	2627	4 US-09-184-445-3	Sequence 3, Appli
10	41.5	43.7	713	2 US-08-849-212-4	Sequence 4, Appli
11	40	42.1	431	1 US-07-783-705A-2	Sequence 1, Appli
12	38.5	40.5	237	2 US-08-578-709-11	Sequence 11, Appli
13	38.5	40.5	345	3 US-09-231-529-1	Sequence 1, Appli
14	38.5	40.5	345	4 US-08-977-816-1	Sequence 15, Appli
15	38.5	40.5	500	2 US-08-578-709-15	Sequence 2, Appli
16	38	40.0	184	1 US-08-353-550-2	Sequence 2, Appli
17	38	40.0	184	2 US-08-551-687-2	Sequence 2, Appli
18	38	40.0	428	1 US-08-353-550-1	Sequence 1, Appli
19	38	40.0	428	2 US-08-551-687-1	Sequence 1, Appli
20	38	40.0	431	3 US-08-807-342B-5	Sequence 5, Appli
21	38	40.0	469	1 US-08-353-550-6	Sequence 6, Appli
22	38	40.0	469	2 US-08-551-687-6	Sequence 6, Appli
23	38	40.0	639	2 US-08-557-309B-37	Sequence 37, Appli
24	38	40.0	639	3 US-08-834-306-37	Sequence 37, Appli
25	38	40.0	639	4 US-08-993-674A-37	Sequence 37, Appli
26	37	38.9	334	2 US-08-566-096A-6	Sequence 6, Appli
27	37	38.9	334	2 US-08-668-650B-6	Sequence 6, Appli

28	37	38.9	334	5 PCT-US95-15646-6	Sequence 6, Appli
29	37	38.9	445	2 US-08-630-118A-2	Sequence 2, Appli
30	37	38.9	445	2 US-08-630-118A-4	Sequence 4, Appli
31	37	38.9	445	2 US-08-838-399-2	Sequence 2, Appli
32	37	38.9	445	2 US-08-838-399-4	Sequence 4, Appli
33	37	38.9	445	2 US-09-003-199-23	Sequence 23, Appli
34	37	38.9	445	4 US-09-235-839-2	Sequence 4, Appli
35	37	38.9	445	4 US-09-235-839-4	Sequence 4, Appli
36	37	38.9	456	1 US-08-349-025-2	Sequence 2, Appli
37	37	38.9	456	2 US-08-566-096A-2	Sequence 2, Appli
38	37	38.9	456	2 US-08-668-650B-2	Sequence 2, Appli
39	37	38.9	456	2 US-08-668-650B-14	Sequence 14, Appli
40	37	38.9	456	5 PCT-US95-15646-2	Sequence 2, Appli
41	36.5	38.4	613	1 US-08-405-615-1	Sequence 1, Appli
42	36.5	38.4	613	2 US-08-461-234-1	Sequence 1, Appli
43	36.5	38.4	613	2 US-08-463-480-1	Sequence 1, Appli
44	36.5	38.4	614	1 US-08-225-224-1	Sequence 1, Appli
45	36.5	38.4	614	3 US-08-722-258-1	Sequence 1, Appli

## ALIGNMENTS

```
RESULT 1
US-09-007-999-2
: Sequence 2, Application US/09007999
: Patent No. 6087559
: GENERAL INFORMATION:
: APPLICANT: Nichols, Scott E.
: TITLE OF INVENTION: Substitutes for Modified Starch and
: FILE REFERENCE: 0356D
: CURRENT APPLICATION NUMBER: US/09/007, 999
: EARLIER FILING DATE: 1998-01-16
: EARLIER APPLICATION NUMBER: 08/478, 704
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 1475
: TYPE: PRT
: ORGANISM: Streptococcus mutans
US-09-007-999-2

Query Match      100.0%; Score 95; DB 3; Length 1475;
Best Local Similarity 100.0%; Pred. No. 9e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 VPSYSFIRAHDSVODLIA 19
Db      552 VPSYSFIRAHDSVODLIA 570

RESULT 2
US-09-210-361-2
: Sequence 2, Application US/09210361
: Patent No. 6284479
: GENERAL INFORMATION:
: APPLICANT: Nichols, Scott E.
: TITLE OF INVENTION: Substitutes for Modified Starches and
: FILE REFERENCE: 0357CR
: CURRENT APPLICATION NUMBER: US/09/210, 361
: EARLIER FILING DATE: 1998-12-11
: EARLIER APPLICATION NUMBER: 09/007, 999
: EARLIER FILING DATE: 1998-01-16
: EARLIER APPLICATION NUMBER: 08/478, 704
: EARLIER FILING DATE: 1995-06-07
: EARLIER APPLICATION NUMBER: 09/009, 620
: EARLIER FILING DATE: 1998-01-20
: EARLIER APPLICATION NUMBER: 08/485, 243
: EARLIER FILING DATE: 1995-06-07
```

```
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-210-361-2
```

```
Query Match          100.0%; Score 95; DB 4; Length 1475;
Best Local Similarity 100.0%; Pred. No. 9e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
1 VPSYSFIRAHDSVODLIA 19
|||||
552 VPSYSFIRAHDSVODLIA 570
```

```
RESULT 3
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-4
```

```
Query Match          95.8%; Score 91; DB 4; Length 1375;
Best Local Similarity 100.0%; Pred. No. 4.4e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 VPSYSFIRAHDSVODLI 18
|||||
DB 578 VPSYSFIRAHDSVODLI 595
```

```
RESULT 4
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
```

```
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2
```

```
Query Match          68.4%; Score 65; DB 3; Length 1430;
Best Local Similarity 76.5%; Pred. No. 0.0026;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 3 SYSFIRAHDSVODLIA 19
|||||
DB 576 NYIFIRAHDSVOTVIA 592
```

```
RESULT 5
US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-6
```

```
Query Match          68.4%; Score 65; DB 4; Length 1430;
Best Local Similarity 76.5%; Pred. No. 0.0026;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 3 SYSFIRAHDSVODLIA 19
|||||
DB 576 NYIFIRAHDSVOTVIA 592
```

```
RESULT 6
US-08-793-824-2
; Sequence 2, Application US/08793824
; Patent No. 5981838
; GENERAL INFORMATION:
; APPLICANT: Simpson, Christine Lynn
; APPLICANT: Giffard, Philip Morrison
; APPLICANT: Jacques, Nicholas Anthony
; TITLE OF INVENTION: Genetic Manipulation of Plants to
; INCREASE STARCH: Increase Stored Carbohydrates
; NUMBER OF SEQUENCES: 2
```



US-09-184-445-3  
; Sequence 3, Application US/09184445  
; Patent No. 6174703  
; GENERAL INFORMATION:  
; APPLICANT: Harrington, Lea A.  
; APPLICANT: Robinson, Murray O.  
; TITLE OF INVENTION: No. 6174703el Genes Encoding Telomerase Protein  
; TITLE OF INVENTION: 1  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen, Inc.  
; STREET: 1840 De Havilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/184,445  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/751,189  
; FILING DATE: 15-NOV-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oleski, Nancy A.  
; REGISTRATION NUMBER: 34,688  
; REFERENCE/DOCKET NUMBER: A-433  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2627 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; US-09-184-445-3

Query Match 44.2%; Score 42; DB 4; Length 2627;  
Best Local Similarity 47.1%; Pred. No. 85;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

1 VPSYSTRAHDSVQDL 17  
:|||| :|||:  
205 MPSYSLSGEEVEDL 221

RESULT 10  
US-08-849-212-4  
; Sequence 4, Application US/08849212  
; Patent No. 5827698  
; GENERAL INFORMATION:  
; APPLICANT: KIKUCHI, YOSHIMI  
; APPLICANT: SUZUKI, TOMOKO  
; APPLICANT: KOTIMA, HIROYUKI  
; TITLE OF INVENTION: NOVEL LYSINE DECARBOXYLASE GENE AND  
; TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,  
; ADDRESSEE: P.C.  
; STREET: 1755 SOUTH JEFFERSON DAVIS HIGHWAY, SUITE 400  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/849,212  
; FILING DATE: 09-JUN-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6/306386  
; FILING DATE: 09-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OBLON, NORMAN F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 10-856-0PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 413-3000  
; TELEFAX: (703) 413-2220  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 713 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-849-212-4

Query Match 43.7%; Score 41.5; DB 2; Length 713;  
Best Local Similarity 45.0%; Pred. No. 22;  
Matches 9; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

QY 1 VPSYSTRAH--DSEVQDL 17  
Db 77 LPLVAFINTHSTMDVSQDM 96

RESULT 11  
US-07-783-705A-2  
; Sequence 2, Application US/07783705A  
; Patent No. 5429939  
; GENERAL INFORMATION:  
; APPLICANT: Misawa, No. 5429939ihiko  
; APPLICANT: Kobayashi, Kazuo  
; APPLICANT: Nakamura, Katsunori  
; APPLICANT: Yamano, Shigeyuki  
; TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE  
; TITLE OF INVENTION: SYNTHESIS OF CAROTENOID  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ladas & Parry  
; STREET: 26 West 61 Street  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10023  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 720kb storage  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: N/A  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/783,705A  
; FILING DATE: 19911023  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 1-103078  
; FILING DATE: 21-APR-1989  
; APPLICATION NUMBER: JP 2-53225  
; FILING DATE: 05-MAR-1990  
; APPLICATION NUMBER: US 07/519,011  
; FILING DATE: 19-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Schwadron, Janet I.  
; REGISTRATION NUMBER: 33,778  
; TELECOMMUNICATION INFORMATION:



;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
;; STREET: 3174 Porter Dr.  
;; CITY: Palo Alto  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 94304  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FASTSEQ for Windows Version 2.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/977,816  
;; FILING DATE: Filed Herewith  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER:  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Billings, Lucy J.  
;; REGISTRATION NUMBER: 36,749  
;; REFERENCE/DOCKET NUMBER: PF-0429 US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 650-855-0555  
;; TELEFAX: 650-845-4166  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 345 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; IMMEDIATE SOURCE:  
;; LIBRARY: KIDNNOT25  
;; CLONE: 353694  
;; US-08-977-816-1

Query Match 40.5%; Score 38.5; DB 4; Length 345;  
Best Local Similarity 40.0%; Pred. No. 32;  
Matches 8; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

QY 1 VPSYSFI--RAHSEVODL 17  
||| : ||| : ||| :  
DB 111 VPSLEKVLKFAHEGIEDL 130

;; JLT 15  
;; 08-578-709-15  
;; Sequence 15, Application US/08578709  
;; Patent No. 5814509  
;; GENERAL INFORMATION:  
;; APPLICANT: TANABE, Tadashi  
;; TITLE OF INVENTION: PROSTACYCLIN SYNTHASE DERIVED FROM HUMAN  
;; NUMBER OF SEQUENCES: 17  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS  
;; STREET: 2100 Pennsylvania Avenue, N.W.  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20037  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/578,709  
;; FILING DATE: 28-DEC-1995  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/JP95/00838

;; FILING DATE: 27-APR-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 114316/1994  
;; FILING DATE: 28-APR-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Gubinsky, Louis  
;; REGISTRATION NUMBER: 24,835  
;; REFERENCE/DOCKET NUMBER: 040439  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202)293-7060  
;; TELEFAX: (202)293-7860  
;; INFORMATION FOR SEQ ID NO: 15:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 500 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-578-709-15

Query Match 40.5%; Score 38.5; DB 2; Length 500;  
Best Local Similarity 29.0%; Pred. No. 50;  
Matches 9; Conservative 4; Mismatches 3; Indels 15; Gaps 1;

QY 3 SYSFI-----RAHSEVODLI 18  
||| : | : ||| :  
DB 168 SYSELRACGLTLTGICETALPRHESQADRV 198

Search completed: March 27, 2002, 13:59:27  
Job time: 580 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:27:00 ; Search time 53.4 Seconds  
(without alignments)  
13.046 Million cell updates/sec

Title: US-09-290-049a-2  
Perfect score: 95  
Sequence: 1 VPSYSTRAHDSVODLIA 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Maximum number of hits satisfying chosen parameters: 100059

Maximum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91	95.8	1375	1	GTFC_STRMU
2	91	95.8	1476	1	GTFB_STRMU
3	86	90.5	1597	1	GTFI_STRDO
4	84	88.4	1592	1	GTF2_STRDO
5	73	76.8	1365	1	GTF5_STRDO
6	65	68.4	1462	1	GTFD_STRMU
7	43	45.3	214	1	COMA_BACSU
8	42	44.2	644	1	RNB_ECOLI
9	42	44.2	659	1	RNB_HAEIN
10	42	44.2	932	1	HIRL_SCHPO
11	41.5	43.7	713	1	DCIL2_ECOLI
12	41	43.2	344	1	MIZD_BACSU
13	40	42.1	205	1	ADEN_ADEG8
14	40	42.1	431	1	CRTX_PANAN
15	40	42.1	548	1	LIP2_CANRU
16	40	42.1	549	1	LIP3_CANRU
17	40	42.1	549	1	LIP4_CANRU
18	40	42.1	572	1	PTLB_STANU
19	40	42.1	6359	1	BACC_BACLI
20	39.5	41.6	192	1	CUPP_DROPS
21	39	41.1	403	1	COAT_BOOLV
22	39	41.1	509	1	CA43_DROME
23	39	41.1	585	1	ASNS_MAIZE
24	39	41.1	715	1	DOAL1_YEAST
25	39	41.1	737	1	SKN1_CANAL
26	39	41.1	946	1	YB16_YEAST
27	39	41.1	1389	1	PAXX_RAT
28	39	41.1	2090	1	N214_HUMAN
29	39	41.1	2116	1	MYS2_DICDI
30	38.5	40.5	454	1	NFM_PLG
31	38.5	40.5	494	1	AINX_HUMAN
32	38.5	40.5	500	1	PGI1_HUMAN
33	38.5	40.5	504	1	AINX_MOUSE

## ALIGNMENTS

RESULT	ID	Query Match	Length	ID	Description
1	GTFC_STRMU	95.8	1375	1	GTFC_STRMU
AC	P13470	95.8	1375	1	GTFC_STRMU
DT	01-NOV-1988	95.8	1375	1	GTFC_STRMU
DT	01-JAN-1990	95.8	1375	1	GTFC_STRMU
DT	15-DEC-1998	95.8	1375	1	GTFC_STRMU
DE	GLUCOSYLTRANSFERASE-SI	95.8	1375	1	GTFC_STRMU
DE	(DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE)	95.8	1375	1	GTFC_STRMU
GN	GTFC	95.8	1375	1	GTFC_STRMU
OS	Streptococcus mutans.	95.8	1375	1	GTFC_STRMU
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;	95.8	1375	1	GTFC_STRMU
OX	NCBI_TaxID=1309;	95.8	1375	1	GTFC_STRMU
RP	SEQUENCE FROM N.A.	95.8	1375	1	GTFC_STRMU
RC	STRAIN-GS-5;	95.8	1375	1	GTFC_STRMU
RX	MEDLINE=89137980; PubMed=2976010;	95.8	1375	1	GTFC_STRMU
RA	Ueda S., Shiroza T., Kuramitsu H.K.;	95.8	1375	1	GTFC_STRMU
RT	"Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.;"	95.8	1375	1	GTFC_STRMU
RL	Gene 69:101-109(1988).	95.8	1375	1	GTFC_STRMU
RN	[2]	95.8	1375	1	GTFC_STRMU
RP	SEQUENCE OF 1-349 FROM N.A.	95.8	1375	1	GTFC_STRMU
RC	STRAIN-GS-5;	95.8	1375	1	GTFC_STRMU
RX	MEDLINE=87308013; PubMed=3040685;	95.8	1375	1	GTFC_STRMU
RA	Shiroza T., Ueda S., Kuramitsu H.K.;	95.8	1375	1	GTFC_STRMU
RT	"Sequence analysis of the gtfB gene from Streptococcus mutans.;"	95.8	1375	1	GTFC_STRMU
RL	J. Bacteriol. 163:4263-4270(1987)	95.8	1375	1	GTFC_STRMU
CC	- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT	95.8	1375	1	GTFC_STRMU
CC	TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE	95.8	1375	1	GTFC_STRMU
CC	OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE	95.8	1375	1	GTFC_STRMU
CC	AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.	95.8	1375	1	GTFC_STRMU
CC	- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) -	95.8	1375	1	GTFC_STRMU
CC	D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).	95.8	1375	1	GTFC_STRMU
CC	- SUBCELLULAR LOCATION: SECRETED.	95.8	1375	1	GTFC_STRMU
CC	- DISEASE: DENTAL CARIES.	95.8	1375	1	GTFC_STRMU
CC	- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA	95.8	1375	1	GTFC_STRMU
CC	1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES). GTF-S SYNTHESIZES	95.8	1375	1	GTFC_STRMU
CC	WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOS). GTF-SI SYNTHESIZES BOTH	95.8	1375	1	GTFC_STRMU
CC	FORMS OF GLUCANS.	95.8	1375	1	GTFC_STRMU
CC	- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-	95.8	1375	1	GTFC_STRMU
CC	BINDING PROTEIN FROM S. MUTANS.	95.8	1375	1	GTFC_STRMU
CC	-----	95.8	1375	1	GTFC_STRMU
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	95.8	1375	1	GTFC_STRMU
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	95.8	1375	1	GTFC_STRMU
CC	the European Bioinformatics Institute. There are no restrictions on its	95.8	1375	1	GTFC_STRMU
CC	use by non-profit institutions as long as its content is in no way	95.8	1375	1	GTFC_STRMU
CC	modified and this statement is not removed. Usage by and for commercial	95.8	1375	1	GTFC_STRMU
CC	entities requires a license agreement (See http://www.isb.ch/announce/	95.8	1375	1	GTFC_STRMU
CC	or send an email to license@isb-sib.ch).	95.8	1375	1	GTFC_STRMU
CC	-----	95.8	1375	1	GTFC_STRMU
DR	EMBL: M22054; AA88592.1; -	95.8	1375	1	GTFC_STRMU
DR	EMBL: M17361; AA88589.1; -	95.8	1375	1	GTFC_STRMU
DR	PIR: J0345; J0345.	95.8	1375	1	GTFC_STRMU
DR	PIR: C3135; C3135.	95.8	1375	1	GTFC_STRMU
DR	InterPro: IPR002479; CW_binding.	95.8	1375	1	GTFC_STRMU

```

DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 7.
DR Pfam: PF0324; Glyco_hydro_70; 1.
KW Transferrase; Glucosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 34
FT CHAIN 35 1375 GLUCOSYLTRANSFERASE-SI.
FT DOMAIN 35 1050 CATALYTIC (APPROXIMATE).
FT DOMAIN 1126 1375 GLUCAN-BINDING (APPROXIMATE).
FT DOMAIN 1126 1375 2.4 A, 1 C AND 1 AC REPEATS.
FT REPEAT 1126 1159 A REPEAT.
FT REPEAT 1169 1200 A REPEAT.
FT REPEAT 1227 1238 C REPEAT.
FT REPEAT 1253 1303 AC REPEAT.
FT REPEAT 1318 1330 A REPEAT (INCOMPLETE).
SQ SEQUENCE 1375 AA; 153022 MW; D4B80CBEE0ACE13 CRC64;

Query Match 95.8%; Score 91; DB 1; Length 1375;
Past Local Similarity 100.0%; Pred. NO. 2.7e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPSYSFIHAESEVDLI 18
| | | | | | | | | | | | | | | | | |
Db 578 VPSYSFIHAESEVDLI 595

RESULT 2
GTFB_STRMU STANDARD: PRT: 1476 AA.
AC P08987; 069381; 069384; 069387; 069390; 069396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)
DE (SUCCROSE 6-GLUCOSYLTRANSFERASE).
GN GTFB.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RX Shiroza T., Ueda S., Kuramitsu H.K.;
RA "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RJ J. Bacteriol. 169:4263-4270(1987).
[2]
SEQUENCE FROM N.A.
RC STRAIN=MT4239, MT4245, MT4251, MT4467, AND MT8148;
RX MEDLINE=98231643; PubMed=9570124;
RX Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
CC -1 FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1 CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) =
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -1 SUBCELLULAR LOCATION: SECRETED.
CC -1 DISEASE: DENTAL CARIES.
CC -1 MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -1 SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

```

CC		the European Bioinformatics Institute. There are no restrictions on its
CC		use by non-profit institutions as long as its content is in no way
CC		modified and this statement is not removed. Usage by and for commercial
CC		entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC		or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC		-----
DR	EMBL; M1761; AAA89588.1; -	
DR	EMBL; D88651; BAA26101.1; -	
DR	EMBL; D88654; BAA26105.1; -	
DR	EMBL; D88657; BAA26109.1; -	
DR	EMBL; D88660; BAA26113.1; -	
DR	EMBL; D89977; BAA26119.1; -	
DR	PIR; B3135; B3135.	
DR	InterPro; IPR002479; CW_binding.70.	
DR	InterPro; IPR003318; Glyco_hydro.70.	
DR	Pfam; PF01473; CW_binding_1; 13.	
DR	Pfam; PF02324; Glyco_hydro.70.1.	
KW	transferase; Glycosyltransferase; Signal; Repeat; Dental caries.	
FT	SIGNAL	1 34
FT	CHAIN	35 1476
FT	DOMAIN	35 1051
FT	REPEAT	1097 1476
FT	DOMAIN	1097 1130
FT	REPEAT	1161 1470
FT	REPEAT	1161 1210
FT	REPEAT	1225 1275
FT	REPEAT	1290 1340
FT	REPEAT	1355 1405
FT	REPEAT	1420 1470
FT	VARIANT	62 62
FT	VARIANT	65 65
FT	VARIANT	68 68
FT	VARIANT	78 78
FT	VARIANT	86 86
FT	VARIANT	89 89
FT	VARIANT	168 168
FT	VARIANT	276 276
FT	VARIANT	399 399
FT	VARIANT	474 474
FT	VARIANT	512 512
FT	VARIANT	519 519
FT	VARIANT	701 701
FT	VARIANT	708 708
FT	VARIANT	938 938
FT	VARIANT	952 957
FT	VARIANT	963 964
FT	VARIANT	968 970
FT	VARIANT	1086 1086
FT	VARIANT	1158 1158
FT	VARIANT	1163 1163
FT	VARIANT	1168 1168
FT	VARIANT	1182 1182
FT	VARIANT	1234 1234
FT	VARIANT	1263 1263
FT	VARIANT	1263 1263
FT	VARIANT	1264 1264
FT	VARIANT	1272 1272
FT	VARIANT	1329 1329
FT	VARIANT	1394 1394
FT	VARIANT	1402 1402
FT	VARIANT	1459 1459
FT	CONFLICT	570 570
FT	CONFLICT	800 817
FT	CONFLICT	1310 1310
SO	SEQUENCE	1476 AA; 165685 MW; 3479B62B07694D98 CRC64;



FT DOMAIN 1093 1592 6.5 X TANDEM REPEATS.  
FT REPEAT 1093 1142 1.  
FT REPEAT 1158 1207 2.  
FT REPEAT 1232 1272 3.  
FT REPEAT 1287 1337 4.  
FT REPEAT 1402 1451 5.  
FT REPEAT 1514 1563 6.  
FT REPEAT 1577 1592 7 (INCOMPLETE).  
SQ SEQUENCE 1592 AA; 176167 MW; BC0A66D079351BECF CRC64;

Query Match 88.4%; Score 84; DB 1; Length 1592;  
Best Local Similarity 88.9%; Pred. No. 4,8e-06;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VPYSFIRAHDSFVODLI 18  
DB 548 VPYSFIRAHDSFVODII 565

RESULT 5  
GTF5\_STRDO STANDARD; PRT: 1365 AA.  
ID GTF5\_STRDO  
AC P29336;  
DT 01-DEC-1992 (Rel. 24, Created)  
DR 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
GN GTF5.  
OS Streptococcus downei (Streptococcus sobrinus).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1317;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MFE28;  
RA MEDLINE=90316665; PubMed=2142479;  
RX Gilmore K.S., Russell R.R., Ferretti J.J.;  
RT "Analysis of the Streptococcus downei gtf5 gene, which specifies a  
glucosyltransferase that synthesizes soluble glucans.";  
RL Infect. Immun. 58:2452-2458(1990).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -  
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
CC -1- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF  
PRIMER GLUCAN UNLIKE GTF-I.  
CC -1- DISEASE: DENTAL CARIES.  
CC -1- MISCELLANEOUS: GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA  
1,6-GLUCOSE).  
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
BINDING PROTEIN FROM S. MUTANS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M30943; AAA26898.1; -  
CC PIR: A41483; A41483.  
CC InterPro: IPR002479; CM\_binding.  
CC InterPro: IPR003318; Glyco\_hydro.70.  
CC Pfam: PF01473; CM\_binding.1; 10.  
CC Pfam: PF02324; Glyco\_hydro.70; 1.  
CC Transfaser: Glycosyltransferase; Signal; Repeat; Dental caries.  
KW SIGNAL  
FT CHAIN 1 36 OR 37 (POTENTIAL).  
FT CHAIN 37 1365 GLUCOSYLTRANSFERASE-S.

FT DOMAIN 37 1050 CATALYTIC (APPROXIMATE).  
FT REPEAT 1083 1365 GLUCAN-BINDING (APPROXIMATE).  
FT DOMAIN 1083 1365 4.5 X TANDEM REPEATS.  
FT REPEAT 1083 1131 1.  
FT REPEAT 1150 1199 2.  
FT REPEAT 1225 1274 3.  
FT REPEAT 1289 1339 4.  
FT REPEAT 1353 1365 5 (INCOMPLETE).  
SQ SEQUENCE 1365 AA; 151590 MW; 16729B5A2E8C476 CRC64;

Query Match 76.8%; Score 73; DB 1; Length 1365;  
Best Local Similarity 78.9%; Pred. No. 0.00029;  
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 VPYSFIRAHDSFVODLIA 19  
DB 537 VPNYVFIKRAHDSFVOTRIA 555

RESULT 6  
GTFD\_STRMU STANDARD; PRT: 1462 AA.  
ID GTFD\_STRMU  
AC P49331; O69383; O69386; O69389; O69392; O69398;  
DT 01-FEB-1996 (Rel. 33, Created)  
DR 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
GN GTFD.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-GS-5;  
RA MEDLINE=91100958; PubMed=2148600;  
RX Honda O., Kato C., Kuramitsu H.K.;  
RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding  
the glucosyltransferase-S enzyme.";  
RL J. Gen. Microbiol. 136:2099-2105(1990).  
CC [2]  
CC SEQUENCE FROM N.A.  
CC STRAIN-MT4239, MT4245, MT4251, MT4467, AND MT8148;  
CC MEDLINE=98231643; PubMed=9570124;  
CC Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
CC Kimura S., Hamada S.;  
CC "Molecular analyses of glucosyltransferase genes among strains of  
CC Streptococcus mutans.";  
CC RT FEWS Microbiol. Lett. 161:331-336(1998).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -  
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- DISEASE: DENTAL CARIES.  
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
CC FORMS OF GLUCANS.  
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
BINDING PROTEIN FROM S. MUTANS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

```

CC EMBL: M29296; AAA26895.1; -
DR EMBL: D88653; BAA26103.1; -
DR EMBL: D88656; BAA26107.1; -
DR EMBL: D88659; BAA26111.1; -
DR EMBL: D88662; BAA26115.1; -
DR EMBL: D89979; BAA26121.1; -
DR InterPro: IPR002479; CW-binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; Glyco_hydro_70.1.
DR Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.
KW SIGNAL.
FT CHAIN 1 1462 GLUCOSYLTRANSFERASE-S.
FT DOMAIN 1232 1423 3 X 63 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 1232 1295 1.
FT REPEAT 1296 1359 2.
FT REPEAT 1360 1423 3.
FT VARIANT 58 58 K -> E (IN STRAIN MT4467).
FT VARIANT 68 68 A -> S (IN STRAIN MT4239 AND MT4245).
FT VARIANT 81 81 T -> T (IN STRAIN MT4251 AND MT8148).
FT VARIANT 113 113 T -> I (IN STRAIN MT4239 AND MT4245).
FT VARIANT 122 122 A -> V (IN STRAIN MT4239, MT4245 AND MT8148).
FT VARIANT 132 132 S -> A (IN STRAIN MT4239, MT4245, MT4251 AND MT8148).
FT VARIANT 135 135 A -> V (IN STRAIN MT4245).
FT VARIANT 202 202 V -> L (IN STRAIN MT4239).
FT VARIANT 255 255 D -> N (IN STRAIN MT8148).
FT VARIANT 275 275 E -> D (IN STRAIN MT4239, MT4245 AND MT4251).
FT VARIANT 288 288 D -> N (IN STRAIN MT4239, MT4245 AND MT4251).
FT VARIANT 301 301 Q -> H (IN STRAIN MT4245).
FT VARIANT 313 313 D -> N (IN STRAIN MT4239 AND MT4251).
FT VARIANT 317 317 V -> K (IN STRAIN MT4239).
FT VARIANT 328 328 E -> F (IN STRAIN MT4239).
FT VARIANT 350 350 F -> L (IN STRAIN MT4239, MT4251 AND MT4467).
FT VARIANT 628 628 KKRIYQ -> EKEVTL (IN STRAIN MT4251).
FT VARIANT 688 688 A -> S (IN STRAIN MT4239).
FT VARIANT 726 726 TDGSGA -> ADKNGS (IN STRAIN MT4251).
FT VARIANT 730 730 TDGGS -> ADKGN (IN STRAIN MT4239 AND MT4245).
FT VARIANT 964 964 D -> Y (IN STRAIN MT4251).
FT VARIANT 1019 1019 E -> K (IN STRAIN MT4245 AND MT4251).
FT VARIANT 1059 1059 LG -> IR (IN STRAIN MT4251).
FT VARIANT 1060 1060 G -> R (IN STRAIN MT4245).
FT VARIANT 1080 1080 G -> R (IN STRAIN MT4239).
FT VARIANT 1142 1142 O -> H (IN STRAIN MT4239, MT4245, MT4251, MT4467 AND MT8148).
FT VARIANT 1198 1198 S -> N (IN STRAIN MT4239).
FT VARIANT 1220 1220 Y -> C (IN STRAIN MT4251 AND MT4467).
FT VARIANT 1280 1280 F -> L (IN STRAIN MT4467).
FT VARIANT 1282 1282 O -> P (IN STRAIN MT4245).
FT VARIANT 1290 1290 K -> T (IN STRAIN MT4245).
FT VARIANT 1311 1311 N -> D (IN STRAIN MT4245).
FT VARIANT 1403 1403 D -> G (IN STRAIN MT4239, MT4245, MT4251 AND MT8148).
FT VARIANT 1425 1425 R -> G (IN STRAIN MT4239, MT4245, MT4251, MT4467 AND MT8148).
FT VARIANT 1449 1449 R -> K (IN STRAIN MT4467).
FT CONFLICT 1428 1462 RYDKNSGNMYNVTLANGRIGIDRWGIARY -> VYR (IN REF. 1).
SO SEQUENCE 1462 AA; 163512 MW; 5C6541F0DCB0DF00 CRC64;

```

Query Match Score 65; DB 1; Length 1462;  
Best Local Similarity 76.5%; Pred. No. 0.007;  
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

3 SYSFIRAHSEYQTVIA 19  
: |||||

```

DB 576 NYFIRAHSEYQTVIA 592
RESULT 7
COMA_BACSU STANDARD: PRT; 214 AA.
ID COMA_BACSU
AC PI4204;
DT 01-JAN-1990 (Rel. 13, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE COMPETENCE PROTEIN A.
GN COMA OR COMAA OR COMA1.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90008771; PubMed=2507523;
RA Weinrauch Y., Guillen N., Dubnau D.;
RT "Sequence and transcription mapping of Bacillus subtilis competence
RT genes comb and coma, one of which is related to a family of bacterial
RT regulatory determinants."
RL J. Bacteriol. 171:5362-5375(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Oudega B., Koningssteijn G., van Zoest A.;
RL Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: REQUIRED FOR THE DEVELOPMENT OF COMPETENCE IN B.
CC SUBTILIS. FOR THE EXPRESSION OF LATE-EXPRESSING COMPETENCE
CC GENES. AND FOR THE EXPRESSION OF THE GROWTH STAGE-REGULATED
CC MOLECULE SURFACTIN. MAY PLAY A REGULATORY ROLE DURING THE
CC DEVELOPMENT OF COMPETENCE, AND AN ANALOGOUS ROLE AS UVRC-ORF2
CC (TRANSDUCING ENVIRONMENTAL INFORMATION TO THE DNA REPAIR SYSTEM).
CC -I- PTM: PHOSPHORYLATED BY COMP.
CC -I- SIMILARITY: TO OTHER BACTERIAL REGULATORY PROTEINS INVOLVED IN
CC SIGNAL TRANSDUCTION.
CC -I- SIMILARITY: BELONGS TO THE LUXR/HRPA FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M2856; AAA22320.1; -
DR EMBL: Z93932; CAB07904.1; -
DR EMBL: Z99120; CAB15156.1; -
DR PIR: A33591; RGSBCA.
DR Subtilisin; BGI0381; COMA.
DR InterPro: IPR000792; HTH_LuxR.
DR InterPro: IPR001789; Response-reg.
DR Pfam: PF00196; Gere; 1.
DR SMART: SMO0421; response_reg; 1.
DR SMART: SMO0448; REC; 1.
DR PROSITE: PS00622; HTH_LuxR_FAMILY; 1.
KW Sensory transduction; Phosphorylation; Transcription regulation;
KW DNA-binding; Activator; Complete proteome.
FT DOMAIN 1 101 RECEIVER DOMAIN.
FT MOD_RES 55 55 PHOSPHORYLATION (BY SIMILARITY).
FT DNA_BIND 171 190 H-T-H MOTIF (BY SIMILARITY).
SO SEQUENCE 214 AA; 24128 MW; 52D01A9740759072 CRC64;

```

Query Match Score 43; DB 1; Length 214;  
Best Local Similarity 52.9%; Pred. No. 4.4;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

```

OY      2 PSYFIRADSEVODLI 18
DB      36 PSYFIRADSEVODLI 52

RESULT 8
RNB_ECOLI STANDARD: PRT: 644 AA.
AC P30850: P78280:
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EXORIBONUCLEASE II (EC 3.1.13.1) (RIBONUCLEASE II) (RNASE II).
GN RNB OR B1286.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
   Escherichia.
NCBI_TaxID=562;
(1)
SEQUENCE FROM N.A.
RC STRAIN-K12:
RA MEDLINE-93268098; PubMed-8497196;
RA Zilhao R., Camelo L., Arraiano C.M.;
RT "DNA sequencing and expression of the gene rnb encoding Escherichia
RT coli ribonuclease II."
RL Mol. Microbiol. 8:43-51(1993).
RN [2]
RP REVISIONS.
RA Zilhao R.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655:
RA MEDLINE-97426617; PubMed-9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-K12:
RA MEDLINE-97251357; PubMed-9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasei H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Samedel G., Seki Y., Sivasubaram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horikuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map."
RL DNA Res. 3:363-377(1996).
RN [5]
RP FUNCTION: INVOLVED IN MRNA DEGRADATION. HYDROLYZES SINGLE-STRANDED
RP POLYRIBONUCLEOTIDES PROCESSIVELY IN THE 3' TO 5' DIRECTION.
RC CATALYTIC ACTIVITY: EXONUCLEOTIC CLEAVAGE IN THE 3' TO 5'
CC DIRECTION TO YIELD 5'-PHOSPHOMONONUCLEOTIDES.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE RIBONUCLEASE II (RNB) FAMILY.
CC -1- SIMILARITY: CONTAINS 1 'SI MOTIF' DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X67913; CAA8112.1;
CC EMBL: AEO00226; AAC74368.1;
CC EMBL: D90767; BAA14848.1;
CC EMBL: D90766; BAA14840.1;

```

```

DR PIR: S28506; S28506.
DR PIR: S32940; S32940.
DR Ecogen; Bg11620; rnb.
DR InterPro: IPR002059; Cold_shock.
DR InterPro: IPR001900; Ribonuclease_II.
DR InterPro: IPR003029; S1.
DR Pfam: PF00773; RNB; 1.
DR Pfam: PF00575; S1; 1.
DR SMART; SM00357; CSP; 1.
DR SMART; SM00316; S1; 1.
DR PROSITE: PS01175; RIBONUCLEASE_II; 1.
DR Hydrolase; Exonuclease; Nuclease; RNA-binding; Complete proteome.
FT CONFLICT 384 384 I -> N (IN REF. 1).
FT CONFLICT 399 399 C -> R (IN REF. 1).
FT CONFLICT 513 513 A -> R (IN REF. 1).
SQ SEQUENCE 644 AA; 72490 MW; 36B16712CDF14394 CRC64;

Query Match 44.2%; Score 42; DB 1; Length 644;
Best Local Similarity 44.4%; Pred. No. 22;
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

OY      2 PSYFIRADSEVODLI 19
DB      362 PSYFIRADSEVODLI 379

RESULT 9
RNB_HAEM
ID RNB_HAEM STANDARD: PRT: 659 AA.
AC P44440;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EXORIBONUCLEASE II (EC 3.1.13.1) (RIBONUCLEASE II) (RNASE II).
GN RNB OR H11733.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
   Haemophilus.
NCBI_TaxID=727;
(1)
RP SEQUENCE FROM N.A.
RC STRAIN-RD / ATCC 51907;
RA MEDLINE-95350630; PubMed-7542800;
RA Fletschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Goeyne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ullrich T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA Fine L.D., Fritschman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
RN [1]
RP FUNCTION: INVOLVED IN MRNA DEGRADATION. HYDROLYZES SINGLE-STRANDED
RP POLYRIBONUCLEOTIDES PROCESSIVELY IN THE 3' TO 5' DIRECTION
CC CATALYTIC ACTIVITY: EXONUCLEOTIC CLEAVAGE IN THE 3' TO 5'
CC DIRECTION TO YIELD 5'-PHOSPHOMONONUCLEOTIDES.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RIBONUCLEASE II (RNB) FAMILY.
CC -1- SIMILARITY: CONTAINS 1 'SI MOTIF' DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X67913; CAA8112.1;
CC EMBL: AEO00226; AAC74368.1;
CC EMBL: D90767; BAA14848.1;
CC EMBL: D90766; BAA14840.1;

```

DR EMBL: U32846; AAC23378.1; -.  
 DR TIGR: H11733; -.  
 DR InterPro: IPR002059; Cold\_shock.  
 DR InterPro: IPR001900; Ribonuclease\_II.  
 DR InterPro: IPR003029; SI.  
 DR Pfam: PF00773; RNB; 1.  
 DR Pfam: PF00575; SI; 1.  
 DR SMART: SM00357; CSP; 1.  
 DR SMART: SM00316; SI; 1.  
 DR PROSITE: PS01175; RIBONUCLEASE\_II; 1.  
 DR Hydrolase; Nuclease; RNA-binding; Complete proteome.  
 KW SEQUENCE 659 AA; 75782 MW; FQADA907A7D3B66 CRC64;

Query Match 44.2%; Score 42; DB 1; Length 659;  
 Best Local Similarity 38.9%; Pred. No. 22;  
 Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 2 PSYFIRAHSEVODLIA 19  
 1  
 367 PDYAFVLAENGKVOEIKR 384

RESULT 10  
 HIRL\_SCHPO STANDARD: PRT: 932 AA.  
 AC H87314; 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE HISTONE TRANSCRIPTION REGULATOR 1 HOMOLOG.  
 GN SPBC31F10.13C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Wood V., Rajandream M.A., Barrell B.G., Pohl T.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: COULD PLAY A PART IN MECHANISMS OF TRANSCRIPTIONAL  
 CC REGULATION SIMILAR TO THAT PLAYED BY YEAST HIR1 AND HIR2 TOGETHER.  
 CC -1- SIMILARITY: BELONGS TO THE HIR1 FAMILY OF WD-REPEAT PROTEINS.  
 CC -1- SIMILARITY: CONTAINS 8 WD REPEATS (TRP-ASP DOMAINS).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: Z97204; CAB10089.1; -.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00400; WD40; 6.  
 DR PRINTS: PR00320; GPROTEINBRPT.  
 DR SMART: SM00320; WD40; 6.  
 DR PROSITE: PS00678; WD\_REPEATS\_1; 1.  
 DR PROSITE: PS50082; WD\_REPEATS\_2; 4.  
 DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
 KW Hypothetical protein; transcription regulation; Repeat; WD repeat;  
 KW Nuclear protein.  
 FT REPEAT 16 55 WD 1.  
 FT REPEAT 72 111 WD 2.  
 FT REPEAT 132 171 WD 3.  
 FT REPEAT 174 213 WD 4.  
 FT REPEAT 268 316 WD 5.  
 FT REPEAT 320 361 WD 6.  
 FT REPEAT 623 663 WD 7.  
 FT REPEAT 664 709 WD 8.

SO SEQUENCE 932 AA; 103686 MW; 09D527CDB9003D3E CRC64;

Query Match 44.2%; Score 42; DB 1; Length 932;  
 Best Local Similarity 46.7%; Pred. No. 32;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 3 SYSFIRAHSEVODL 17  
 1  
 DB 126 SYRLLGHNDNDIQLD 140

RESULT 11  
 DCLZ\_ECOLI STANDARD: PRT: 713 AA.  
 AC P52095; P78299; 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE LYSINE DECARBOXYLASE, CONSTITUTIVE (EC 4.1.1.18) (LDC).  
 GN LDC OR LDC OR LDC OR B0186.  
 GN LDC OR LDC OR LDC OR B0186.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.  
 RC STRAIN=K12 / W3110;  
 RX MEDLINE=97369816; PubMed=9226257;  
 RA Kikuchi Y., Kojima H., Tanaka T., Takatsuka Y., Kamio Y.;  
 RT "Characterization of a second lysine decarboxylase isolated from  
 RT Escherichia coli.";  
 RL J. Bacteriol. 179:4486-4492(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-40 AND 415-456.  
 RC STRAIN=K12 / W3110;  
 RX MEDLINE=97480927; PubMed=9339543;  
 RA Yamamoto Y., Miya Y., Miyoshi K., Furuyama J., Ohmori H.;  
 RT "The Escherichia coli ldc gene encodes another lysine decarboxylase,  
 RT probably a constitutive enzyme.";  
 RL Genes Genet. Syst. 72:167-172(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.R., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Snares V.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1233-1238(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / W3110;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,  
 RA Mizuno T., Takino K., Nakata A., Yura T., Sampei G., Mizobuchi K.;  
 RT "Systematic sequencing of the Escherichia coli genome: analysis of the  
 RT 4.0 - 6.0 min (189,987 - 281,416bp) region.";  
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=96389317; PubMed=9723924;  
 RA Pichoff S., Alibaud L., Guedant A., Castanie M.-P., Bouche J.-P.;  
 RT "An Escherichia coli gene (yaeO) suppresses temperature-sensitive

RT mutations in essential genes by modulating Rho-dependent  
 RT transcription termination."  
 RL Mol. Microbiol. 29:859-869(1998).  
 CC -1- FUNCTION: LDC IS CONSTITUTIVELY BUT WEAKLY EXPRESSED UNDER VARIOUS  
 CC CONDITIONS. OPTIMUM ACTIVITY IS ACHIEVED BETWEEN PH 6.2 TO 8.0.  
 CC -1- CATALYTIC ACTIVITY: L-LYSINE - CADAVERINE + CO(2).  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -1- SUBUNIT: HOMODECAMER.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF ORNITHINE, LYSINE, AND ARGININE  
 CC DECARBOXYLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: D87518; BAA21656.1; -  
 DR EMBL: D49445; BAA08426.1; -  
 DR EMBL: AE000128; AAC73297.1; -  
 DR EMBL: D83536; BAA77861.1; -  
 DR EMBL: U70214; AAB08615.1; -  
 DR EMBL: Z50870; CAA90749.1; -  
 DR Ecogene: EG13219; ldcC.  
 DR InterPro: IPR000310; Orn\_Lys\_Arg\_decarboxylase\_1.  
 DR Pfam: PF01276; OKR\_DC\_1; 1.  
 DR PROSITE: PS00703; OKR\_DC\_1; 1.  
 KW Lyase; Decarboxylase; Pyridoxal phosphate; Multigene family;  
 KW Complete proteome.  
 FT BINDING 367 PYRIDOXAL PHOSPHATE.  
 FT CONFLICT 284 S -> T (IN REF. 2).  
 FT CONFLICT 314 N -> F (IN REF. 2).  
 FT CONFLICT 411 T -> S (IN REF. 2).  
 FT CONFLICT 413 AA -> R (IN REF. 2).  
 FT CONFLICT 438 MA -> I (IN REF. 2).  
 FT CONFLICT 673 L -> I (IN REF. 2).  
 SQ SEQUENCE 713 AA; 80590 MW; 4532C6069744ABDF CRC64;

Query Match 43.7%; Score 41.5; DB 1; Length 713;  
 Best Local Similarity 45.0%; Pred. No. 29;  
 Matches 9; Conservative 3; Mismatches 5; Indels 3; Gaps 1;  
 QY 1 VPYSFIRAH---DSEVQDL 17  
 77 LPLYAFINTHSTMDVSVDQM 96

RESULT 12  
 M12D\_BACSU STANDARD: PRT; 344 AA.  
 AC P26935;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE MYO-INOSITOL 2-DEHYDROGENASE (EC 1.1.1.18).  
 GN IDH OR IOLG OR E83G.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=60015;  
 RX MEDLINE=92104493; PubMed=1761221;  
 RA Fujita Y., Shindo K., Miwa Y., Yoshida K.;  
 RT "Bacillus subtilis inositol dehydrogenase-encoding gene (idh):  
 RT sequence and expression in Escherichia coli.";  
 RL Gene 108:121-125(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC STRAIN=168 / BGSC1A1;  
 RX MEDLINE=95039891; PubMed=7952181;  
 RA Yoshida K., Sano H., Miwa Y., Ogasawara N., Fujita Y.;  
 RT "Cloning and nucleotide sequencing of a 15 kb region of the Bacillus  
 RT subtilis genome containing the iol operon.";  
 RL Microbiology 140:2289-2298(1994).  
 RN [3]  
 RP CHARACTERIZATION.  
 RX MEDLINE=79239346; PubMed=112095;  
 RA Ramaley R., Fujita Y., Freese E.;  
 RT "Purification and properties of Bacillus subtilis inositol  
 RT dehydrogenase.";  
 RL J. Biol. Chem. 254:7684-7690(1979).  
 CC -1- CATALYTIC ACTIVITY: MYO-INOSITOL + NAD(+) -> 2,4,6/3,5-  
 CC PENTAHYDROXYCYCLOHEXANONE + NADH.  
 CC -1- PATHWAY: FIRST STEP OF MYO-INOSITOL CATABOLISM.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- INDUCTION: INOSITOL, SUBJECTED TO CATABOLITE REPRESSION.  
 CC -1- SIMILARITY: BELONGS TO THE GFO/IDH/MOCA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M76431; AAA22543.1; -  
 DR EMBL: D14399; BAA03296.1; -  
 DR EMBL: Z99124; CAB16006.1; -  
 DR PIR: JH0511; JH0511.  
 DR Subtilist: BG10669; idh.  
 DR InterPro: IPR000683; GFO\_IDH\_MOCA.  
 DR Pfam: PF01408; GFO\_IDH\_MOCA; 1.  
 KW Oxidoreductase; NAD; Complete proteome.  
 SQ SEQUENCE 344 AA; 38351 MW; 2FCE908DAE2C332P CRC64;

Query Match 43.2%; Score 41; DB 1; Length 344;  
 Best Local Similarity 53.8%; Pred. No. 16;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 6 FIRAHSEVQDLI 18  
 Db 276 EVAADVEIQDFI 288

RESULT 13  
 ADEN\_ADEG8 STANDARD: PRT; 205 AA.  
 AC Q90W72;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ADENAIN (EC 3.4.22.39) (ENDOPROTEASE) (LATE L3 23 KDA PROTEIN).  
 OS Avian adenovirus type 8 (strain AFCC A-2A) (Fowl adenovirus 8).  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadnavirus.  
 OX NCBI\_TaxID=66295;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ojic D., Nagy E.;  
 RT "The DNA sequence of fowl adenovirus 8.";  
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBU databases.  
 CC -1- FUNCTION: THIOLESTERASE CLEAVING SPECIFIC GLY-ALA PEPTIDES IN A  
 CC NUMBER OF VIRAL PRECURSOR PROTEINS (TERMINAL, IIA, VI, VII, VIII,  
 CC IIA AND MO) WHICH ARE REQUIRED FOR VIRION MATURATION. ALSO CLEAVES  
 CC HOST CELLS CYTOSKELETAL KERATINS K7 AND K18.  
 CC -1- CATALYTIC ACTIVITY: CLEAVES PROTEINS OF THE ADENOVIRUS AND ITS  
 CC HOST CELL AT TWO CONSENSUS SITES: -YAA-XAA-GLY-GLY-I-XAA-AND -YAA-  
 CC XAA-GLY-XAA-I-GLY-(IN WHICH YAA IS MET, ILE OR LEU, AND XAA IS ANY  
 CC AMINO ACID).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C5.



```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC / between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF083975; MAD50345.2; -
DR HSSP; P03252; IAP.
DR MEROPS; C05.001; -.
DR InterPro; IPR000855; Peptidase_C5.
DR Pfam; PF00770; Peptidase_C5; 1.
DR PRINTS; PR00703; ADYENDOPPEASE.
DR ProDom; PD003705; Peptidase_C5; 1.
DR HydroLase; Thiol protease; Late protein.
FT ACT_SITE 55 55 BY SIMILARITY.
FT ACT_SITE 72 72 BY SIMILARITY.
FT ACT_SITE 122 122 BY SIMILARITY.
SEQUENCE 205 AA; 23701 MW; 36F0700CDFB85F62 CRC64;

Query Match 42.1%; Score 40; DB 1; Length 205;
Best Local Similarity 63.6%; Pred. No. 13;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 5 SFIRAHSEVQ 15
DB 183 SFFRAHESELK 193

RESULT 14
CRTX_PANAN STANDARD; PRT; 431 AA.
AC P21686;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ZEXANTHIN GLUCOSYL TRANSFERASE (EC 2.4.1.-).
GN CRTX OR UGT101.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=20D3;
RX MEDLINE=91072214; PubMed=2254247;
Misawa N., Nakagawa M., Kobayashi K., Yamano S., Izawa Y.,
Nakamura K., Harashina K.;
"Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway
by functional analysis of gene products expressed in Escherichia
coli";
RT J. Bacteriol. 172:6704-6712(1990).
RL -I- FUNCTION: CATALYZES THE GLYCOSYLATION REACTION WHICH CONVERTS
ZEXANTHIN TO ZEXANTHIN-BETA-DIGLUCOSIDE.
CC -I- PATHWAY: CAROTENOID BIOSYNTHESIS.
CC -I- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; D90087; BAA14125.1; -
DR PIR; B37802; B37802.
DR InterPro; IPR002213; UDPGT.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase; Carotenoid biosynthesis.

```

```

SQ SEQUENCE 431 AA; 47241 MW; 9EC27A798608BC9D CRC64;

Query Match 42.1%; Score 40; DB 1; Length 431;
Best Local Similarity 44.4%; Pred. No. 30;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 2 PSYSFRADHSEVODLIA 19
DB 10 PFYSHVKALQNLQELVA 27

RESULT 15
LIP2_CANRU STANDARD; PRT; 548 AA.
ID P32946;
AC P32946;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE LIPASE 2 PRECURSOR (EC 3.1.1.3).
GN LIP2.
OS Candida rugosa (Yeast) (Candida cylindracea).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5481;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14830;
RX MEDLINE=92305068; PubMed=1610906;
Longhi S., Fusetti F., Grandori R., Lotli M., Vanoni M.,
Alberghina L.;
"Cloning and nucleotide sequences of two lipase genes from Candida
cylindracea";
RT Biochim. Biophys. Acta 1131:227-232(1992).
RL [2]
RN [2]
RP REVIEW.
RX MEDLINE=98451816; PubMed=9778794;
Benjamin S., Pandey A.;
"Candida rugosa lipases: molecular biology and versatility in
biotechnology.";
RT Yeast 14:1069-1087(1998).
RL -I- CATALYTIC ACTIVITY: TRIACYLGLYCEROL + H(2)O = DIACYLGLYCEROL +
A FATTY ACID ANION.
CC -I- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLLESTERASE/LIPASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X64704; CAA45958.1; -
DR PIR; S32615; S32615.
DR HSSP; P32947; LCLE.
DR InterPro; IPR002018; Carboxylesterase_B.
DR InterPro; IPR000379; Est_lip_thioest_actsite.
DR Pfam; PF00135; Coesterase; 1.
DR PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.
DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.
KW HydroLase; Lipid degradation; Signal; Glycoprotein; Multigene family.
FT SIGNAL 1 14
FT CHAIN 15 548 LIPASE 2.
FT ACT_SITE 223 223 BY SIMILARITY.
FT ACT_SITE 355 355 BY SIMILARITY.
FT ACT_SITE 463 463 BY SIMILARITY.
FT DISULFID 74 111 BY SIMILARITY.
FT DISULFID 282 291 BY SIMILARITY.
FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQUENCE 548 AA; 58864 MW; E0DBCFE2501E7614 CRC64;

```

Query Match 42.1%; Score 40; DB 1; Length 548;  
Best Local Similarity 53.3%; Pred. No. 39;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 5 SFRHDSSEVQDLIA 19  
||| | | : | : |  
Db 379 SFRHSDAEIDTLMA 393

Search completed: March 27, 2002, 14:27:01  
Job time: 1643 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:26:01 ; Search time 188.53 Seconds  
(without alignments)  
14.741 Million cell updates/sec

Title: US-09-290-049a-2  
Perfect score: 95  
Sequence: 1 VPSYFIRAHSEVODLIA 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: SPREMBL\_17:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.podent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91	95.8	1390	2	069385 streptococc
2	91	95.8	1455	2	069382 streptococc
3	91	95.8	1455	2	069388 streptococc
4	91	95.8	1455	2	069391 streptococc
5	91	95.8	1455	2	069397 streptococc
6	84	88.4	1590	2	055263 streptococc
7	84	88.4	1590	2	059983 streptococc
8	80	84.2	1477	2	091466 leuconostoc
9	80	84.2	1508	2	052224 leuconostoc
10	80	84.2	1508	2	09EZH5 leuconostoc
11	80	84.2	1527	2	09ZAR4 leuconostoc
12	79	83.2	1290	2	048756 leuconostoc
13	77	81.1	1016	2	091CJ7 leuconostoc
14	74	77.9	1512	2	09WJ5 streptococc
15	73	76.8	1338	2	09WJ4 streptococc
16	66	69.5	2057	2	09RE05 leuconostoc
17	65	68.4	1449	2	068542 streptococc
18	65	68.4	1449	2	055264 streptococc
19	65	68.4	1518	2	Q00600 streptococc

20	64	67.4	1575	2	09LCH3 streptococc
21	64	67.4	1577	2	054178 streptococc
22	64	67.4	1599	2	000599 streptococc
23	61	64.2	1577	2	055265 streptococc
24	50	52.6	93	2	092IX9 streptococc
25	46	48.4	66	2	09K5K6 bacillus mo
26	46	48.4	575	5	P90900 caenorhabdi
27	46	48.4	1032	10	09FIC9 arabidopsis
28	44	46.3	327	10	09C7I1 arabidopsis
29	44	46.3	371	8	09B8A8 arabidopsis
30	43	45.3	51	2	09RNP7 trichinella
31	43	45.3	64	2	09K5L0 bacillus su
32	43	45.3	1260	10	09LMD9 arabidopsis
33	42	44.2	247	10	09M8T9 arabidopsis
34	42	44.2	431	2	047843 enterobacte
35	42	44.2	563	2	09JZL7 neisseria m
36	42	44.2	563	2	09JUP8 neisseria m
37	42	44.2	2627	4	099973 homo sapien
38	41.5	43.7	1882	3	09P898 podospira a
39	41	43.2	145	1	027766 methanobact
40	41	43.2	357	2	033169 mycobacteri
41	41	43.2	567	12	09DHS5 yaba-like d
42	41	43.2	597	2	09PR58 ureaplasma
43	41	43.2	658	2	09CP74 pasteurella
44	41	43.2	738	4	09P2P8 homo sapien
45	41	43.2	1146	2	051778 borrelia bu

## ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT; 1390 AA.
ID 069385		
AC 069385;		
DT 01-AUG-1998 (TREMBLrel. 07, Created)		
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)		
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE GLUCOSYLTRANSFERASE-SI.		
GN GTF.		
OS Streptococcus mutans.		
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		
OC Streptococcus.		
OX NCBI_TaxID=1309;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC STRAIN=MT4245;		
RA MEDLINE=98231643; PubMed=9570124;		
RX Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S., Kimura S., Hamada S.,		
RT "Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans."		
RL EMBL D88633; FEA26106.1; "		
DR InterPro: IPR002479; Glyco_hydro_70.		
DR InterPro: IPR003318; Glyco_hydro_70.		
DR Pfam: PF01473; CW_binding_1; 7.		
DR Pfam: PF02324; Glyco_hydro_70; 1.		
DR TRANSFERASE.		
SQ SEQUENCE 1390 AA; 155375 MW; 8847E4956FE05E9F CRC64;		
Query Match	95.8%;	Score 91; DB 2; Length 1390;
Best Local Similarity	100.0%;	Pred. No. 8.3e-07;
Matches 18; Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Oy 1 VPSYFIRAHSEVODLI 18		
Db 578 VPSYFIRAHSEVODLI 595		
RESULT 2		
069382		

ID 069382 PRELIMINARY; PRT; 1455 AA.  
AC 069382;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTF.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT8148;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S., Kimura S., Hamada S.;  
"Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans.";  
FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL; D88652; BAA26102.1; -;  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 9.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1455 AA; 162969 MW; 27DAD3A1EECA2939 CRC64;

Query Match 95.8%; Score 91; DB 2; Length 1455;  
Best Local Similarity 100.0%; Pred. No. 9e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHDSVODLI 18  
|||||  
DB 578 VPSYSFIRAHDSVODLI 595

RESULT 3  
069388 PRELIMINARY; PRT; 1455 AA.  
AC 069388;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTF.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT4239;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S., Kimura S., Hamada S.;  
"Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans.";  
FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL; D88658; BAA26110.1; -;  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 10.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1455 AA; 163045 MW; 6D90A4978D35DD82 CRC64;

Query Match 95.8%; Score 91; DB 2; Length 1455;  
Best Local Similarity 100.0%; Pred. No. 9e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHDSVODLI 18  
|||||  
DB 578 VPSYSFIRAHDSVODLI 595

RESULT 4  
069391 PRELIMINARY; PRT; 1455 AA.  
AC 069391;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTF.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT4251;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S., Kimura S., Hamada S.;  
"Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans.";  
FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL; D88661; BAA26114.1; -;  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 9.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SQ SEQUENCE 1455 AA; 162804 MW; 683A359D873E9E1A CRC64;

Query Match 95.8%; Score 91; DB 2; Length 1455;  
Best Local Similarity 100.0%; Pred. No. 9e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHDSVODLI 18  
|||||  
DB 578 VPSYSFIRAHDSVODLI 595

RESULT 5  
069397 PRELIMINARY; PRT; 1455 AA.  
AC 069397;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTF.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT4467;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S., Kimura S., Hamada S.;  
"Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans.";  
FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL; D89978; BAA26120.1; -;  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 9.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.

SQ SEQUENCE 1455 AA; 162913 MW; A1263427BF24EBE1 CRC64;

Query Match 95.8%; Score 91; DB 2; Length 1455;

Best Local Similarity 100.0%; Pred. No. 9e-07; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHSEVODLI 18  
|||||  
DB 578 VPSYSFIRAHSEVODLI 595

RESULT 6

055263 PRELIMINARY; PRT; 1590 AA.

AC 055263;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE GTF-I.  
Streptococcus sobrinus.

Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
Streptococcus.

NCBI\_TaxID=1310;

SEQUENCE FROM N.A.  
STRAIN-ATCC 33478;

Sato S.;  
Ann. Kagoshima Univ. Dental School 16:23-29(1996).

EMBL; D63570; BAA09792.1; -;  
InterPro: IPR002479; CW\_binding.

InterPro: IPR003318; Glyco\_hydro\_70.

Pfam: PF01473; CW\_binding\_1; 15.

Pfam: PF02324; Glyco\_hydro\_70; 1.

TRANSFERASE.

SEQUENCE 1590 AA; 176057 MW; 9DF7A3F2C6E4FD43 CRC64;

Query Match 88.4%; Score 84; DB 2; Length 1590;

Best Local Similarity 88.9%; Pred. No. 1.6e-05; Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHSEVODLI 18  
|||||  
DB 548 VPSYSFIRAHSEVODII 565

RESULT 7

059983 PRELIMINARY; PRT; 1590 AA.

AC 059983;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).  
GTF-I.  
Streptococcus sobrinus.

Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
Streptococcus.

NCBI\_TaxID=1310;

SEQUENCE FROM N.A.  
STRAIN-OM2176;

MELINE-94146405; PubMed-8312602;

Sato S.; Inoue M.; Handa N.; Aizawa Y.; Isebe Y.; Katayama T.;

"DNA sequence of the glucosyltransferase gene of serotype d  
Streptococcus sobrinus.";

DNA Seq. 4:19-27(1993).

-I- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) - D-  
FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).

EMBL; D13858; BAA02976.1; -;  
InterPro: IPR002479; CW\_binding.

DR InterPro: IPR003318; Glyco\_hydro\_70.

DR Pfam: PF01473; CW\_binding\_1; 16.

DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW Signal: transferase; Glycosyltransferase.

FT SIGNAL 1  
CHAIN 39 1590 POTENTIAL.  
SEQUENCE 1590 AA; 175955 MW; C3C83A57CF3C2B0E CRC64;

Query Match 88.4%; Score 84; DB 2; Length 1590;

Best Local Similarity 88.9%; Pred. No. 1.6e-05; Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHSEVODLI 18  
|||||  
DB 548 VPSYSFIRAHSEVODII 565

RESULT 8

091466 PRELIMINARY; PRT; 1477 AA.

AC 091466;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE DEXTRANSUCRASE (EC 2.4.1.5).

DSRC.  
Leuconostoc mesenteroides.

Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
Leuconostoc.

NCBI\_TaxID=1245;

SEQUENCE FROM N.A.  
STRAIN-NRL B-1355;

Arguello-Morales M.A.; Remaud-Simeon M.; Pizut S.; Sarcabal P.;

Willmot R.M.; Monsan P.;

"Sequence analysis of the gene encoding alternansucrase, a sucrose  
RT glucosyltransferase from Leuconostoc mesenteroides NRL B-1355.";

Submitted (CCF-1999) to the EMBL/Genbank/DBJ databases.

EMBL; AJ250172; CAB76565.1; -;

InterPro: IPR002479; CW\_binding.

InterPro: IPR003318; Glyco\_hydro\_70.

Pfam: PF01473; CW\_binding\_1; 14.

Pfam: PF02324; Glyco\_hydro\_70; 1.

TRANSFERASE; Glycosyltransferase.

SEQUENCE 1477 AA; 164886 MW; E6F5710DEDFC8B31 CRC64;

Query Match 84.2%; Score 80; DB 2; Length 1477;

Best Local Similarity 73.7%; Pred. No. 7.1e-05; Matches 14; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHSEVODLIA 19  
:|||||  
DB 603 IPNYSFVRHSEVQTVIA 621

RESULT 9

052224 PRELIMINARY; PRT; 1508 AA.

AC 052224;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE 6-  
GLUCOSYLTRANSFERASE).  
DSRB.  
Leuconostoc mesenteroides.

Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
Leuconostoc.

NCBI\_TaxID=1245;

SEQUENCE FROM N.A.

RC STRAIN-NRRL B-1299;  
 RA Monchois V., Remaud-Simeon M., Monsan P., Willemot R.M.,  
 RL FEMS Microbiol. Lett. 0:0-0(1998).  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-  
 FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).  
 DR EMBL: AF030129; AAB95453.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 14.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 DR Transferrase; Glycosyltransferase.  
 KW SEQUENCE 1508 AA; 168511 MW; E70CECB57A70D1F0 CRC64;

Query Match 84.2%; Score 80; DB 2; Length 1508;  
 Best Local Similarity 73.7%; Pred. No. 7.2e-05;  
 Matches 14; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

1 VPSYFIRAHSEVODLIA 19  
 :|:|||||:|||||:|  
 DB 634 IPNYSFVRAHSEVQTVIA 652

RESULT 10  
 O9EZHS PRELIMINARY; PRT; 1508 AA.  
 AC O9EZHS:  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE DSRB742.  
 GN DSRB742.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 CC Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-B-742CB;  
 RA Kim H.-S., Kim D., Ryu H.-J., Robyt J.F.;  
 RT "Leuconostoc mesenteroides B-742CB, a dextranucrase gene";  
 RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF294469; AAG38021.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 14.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 SEQUENCE 1508 AA; 168542 MW; E2FCFA0F87AE4F3A CRC64;

Query Match 84.2%; Score 80; DB 2; Length 1508;  
 Best Local Similarity 73.7%; Pred. No. 7.2e-05;  
 Matches 14; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 VPSYFIRAHSEVODLIA 19  
 :|:|||||:|||||:|  
 DB 634 IPNYSFVRAHSEVQTVIA 652

RESULT 11  
 O9ZAR4 PRELIMINARY; PRT; 1527 AA.  
 AC O9ZAR4:  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE.  
 GN DEX.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 CC Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN (1)

RP SEQUENCE FROM N.A.  
 RC STRAIN-NRRL B-512-F;  
 RA Bhattacharjee R., Singh D.K.S.;  
 RT "Cloning and Molecular Characterization of Dextranucrase Gene from  
 Leuconostoc mesenteroides NRRL B-512F.";  
 RL Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: U81374; AAD10952.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 16.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW SEQUENCE 1527 AA; 169709 MW; 1DFAFA237C743398 CRC64;

Query Match 84.2%; Score 80; DB 2; Length 1527;  
 Best Local Similarity 73.7%; Pred. No. 7.3e-05;  
 Matches 14; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 VPSYFIRAHSEVODLIA 19  
 :|:|||||:|||||:|  
 DB 652 IPNYSFVRAHSEVQTVIA 670

RESULT 12  
 O48756 PRELIMINARY; PRT; 1290 AA.  
 AC O48756:  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 CC Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NRRL B1299;  
 RX MEDLINE-97136686; PubMed-8982063;  
 RA Monchois V., Willemot R.M., Remaud-Simeon M., Croux C., Monsan P.;  
 RT "Cloning and sequencing of a gene coding for a novel dextranucrase  
 from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-  
 6) and alpha (1-3) linkages.";  
 RL Gene 182:23-32(1996).  
 DR EMBL: U38181; AAB40875.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 11.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 SEQUENCE 1290 AA; 145590 MW; 3555C2E96B749F9A CRC64;

Query Match 83.2%; Score 79; DB 2; Length 1290;  
 Best Local Similarity 83.3%; Pred. No. 9e-05;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 PYSYFIRAHSEVODLIA 19  
 :|:|||||:|||||:|  
 DB 388 IPNYSFVRAHSEVQTVIA 405

RESULT 13  
 O9LCJ7 PRELIMINARY; PRT; 1016 AA.  
 AC O9LCJ7:  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE.  
 GN DSRT.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 CC Leuconostoc.

OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B-512F;  
 RA MEDLINE=20169623; PubMed=10705445;  
 RA Funae K., Mizuno K., Takahara H., Kobayashi M.;  
 RT "Gene encoding a dextranucrase-like protein in *Leuconostoc mesenteroides* NRRL B-512F.";  
 RL Biosci. Biotechnol. Biochem. 64:29-38(2000).  
 DR EMBL; AB020020; BAA90527.1; -;  
 DR InterPro; IPR003318; Glyco\_hydro\_70.1.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1016 AA; 110343 MW; 8896FEDE13CCB47 CRC64;

Query Match 81.1%; Score 77; DB 2; Length 1016;  
 Best Local Similarity 77.8%; Pred. No. 0.00015;  
 Matches 14; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 2 PSYSFIRAHDESEVODLIA 19  
 |||:|||||||:11  
 625 PNYVTFIRAHDESEVOTVIA 642

RESULT 14  
 O9WXJ5 PRELIMINARY; PRT; 1512 AA.  
 AC O9WXJ5; 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE GTF-S.  
 GN GTFP.  
 OS Streptococcus criceti.  
 OC Plasmid PAM1.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1333;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HS-6;  
 RA Inoue M., Fukui K., Miyagi A.;  
 RT "S. cricetus glucosyltransferase(gtfs and gftf) genes.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB026123; BAA77237.1; -;  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 14.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR Plasmid.  
 SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9C6C601FC14 CRC64;

Query Match 77.9%; Score 74; DB 2; Length 1512;  
 Best Local Similarity 77.8%; Pred. No. 0.00078;  
 Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 PSYSFIRAHDESEVODLIA 19  
 |||:|||||||:11  
 560 PSYVVFIRAHDESEVOTVIA 577

RESULT 15  
 O9WXJ4 PRELIMINARY; PRT; 1338 AA.  
 AC O9WXJ4; 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE GTF-S.  
 GN GTFP.  
 OS Streptococcus criceti.  
 OC Plasmid PAM1.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1333;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HS-6;  
 RA Inoue M., Fukui K., Miyagi A.;  
 RT "S. cricetus glucosyltransferase(gtfs and gftf) genes.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB026123; BAA77236.1; -;  
 DR HSSP; P06278; IVUS.  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 10.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR Plasmid.  
 SQ SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;

Query Match 76.8%; Score 73; DB 2; Length 1338;  
 Best Local Similarity 76.9%; Pred. No. 0.001;  
 Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 VPSYFIRAHDESEVODLIA 19  
 |||:|||||||:11  
 509 VPNVTFIRAHDESEVOTRIA 527

Search completed: March 27, 2002, 14:26:02  
 Job time: 1674 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:57:53 ; Search time 198.55 Seconds  
(without alignments)  
8.208 Million cell updates/sec

Title: US-09-290-049a-3

Perfect score: 112

Sequence: 1 TGAATNGQLLYFRANGVQVKG 22

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues 522463

Maximum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A-Geneseq\_1101:\*

1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:\*

2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:\*

3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:\*

4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:\*

5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:\*

6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:\*

7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:\*

8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:\*

9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:\*

10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:\*

11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:\*

12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:\*

13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:\*

14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:\*

15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:\*

16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:\*

17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:\*

18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:\*

19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:\*

20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:\*

21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:\*

22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87	77.7	22	AA43696	GT subsequence. S
2	87	77.7	22	AAW34159	GT antigenic pept
3	87	77.7	22	AAV4338	GT antigenic epit
4	87	77.7	24	AAW34164	GT antigenic pept
5	87	77.7	1592	AA32925	Glucosyltransferase
6	66	58.9	1577	AA891047	Alpha-D-glucosyltr
7	65	58.0	2057	AA810667	L. mesenteroides a
8	49.5	44.2	320	AAW87969	Toxin A immunogeni
9	49.5	44.2	457	AAW87970	Toxin A immunogeni
10	49.5	44.2	811	AA895014	C. difficile toxin
11	49.5	44.2	812	AA895017	C. difficile toxin

12	49.5	44.2	862	20	AAW87971	Toxin A immunogeni
13	49.5	44.2	866	21	AA830543	A recombinant prot
14	49.5	44.2	866	21	AA830546	A recombinant prot
15	49.5	44.2	2710	17	AA895016	C. difficile toxin
16	49.5	44.2	2710	19	AAW68387	Clostridium diffic
17	49	43.8	337	22	AA845695	P. faeciparum yfgb
18	47	42.0	785	22	AAW39393	Human polypeptide
19	47	42.0	794	22	AAW41179	Human polypeptide
20	46	41.1	619	15	AA863437	Pneumococcal surfa
21	46	41.1	619	17	AA887598	Pneumococcal surfa
22	46	41.1	619	17	AA886911	Pneumococcal surfa
23	46	41.1	641	20	AAV41838	Streptococcus pneu
24	46	41.1	641	19	AAW61217	Streptococcus pneu
25	46	41.1	648	19	AAW70336	Pneumococcal surfa
26	46	41.1	648	19	AAW62274	Streptococcus pneu
27	46	41.1	648	20	AAV41837	Streptococcus pneu
28	46	41.1	648	20	AAW87879	A pneumococcal sur
29	46	41.1	653	13	AA827150	Pept. fragment. St
30	46	41.1	653	20	AAW92456	S. pneumoniae trun
31	46	41.1	684	16	AA873912	Streptococcus pneu
32	46	41.1	745	21	AA81652	Streptococcus pneu
33	46	41.1	1301	13	AA828304	Acid alpha-amylase
34	45	40.2	204	20	AAW87968	Toxin A immunogeni
35	45	40.2	563	21	AA831199	Arabidopsis thalia
36	45	40.2	630	21	AA831198	Arabidopsis thalia
37	44	39.3	94	21	AA800949	Human secreted pro
38	44	39.3	121	21	AAV54410	Protein encoded by
39	44	39.3	181	21	AAV54411	Amino acid sequenc
40	43.5	38.8	196	22	AA890590	C glutamicum prote
41	43	38.4	153	21	AAV71284	Streptococcus pneu
42	43	38.4	349	22	AAW39747	Human polypeptide
43	43	38.4	380	22	AAW41533	Human polypeptide
44	43	38.4	1036	21	AA801849	Hemophilus Influe
45	43	38.4	1477	14	AA841724	High molecular wei

#### ALIGNMENTS

RESULT 1

ID AAR43696 standard; peptide: 22 AA.

AC AAR43696;

XX 20-MAY-1994 (first entry)

DE GT subsequence.

XX

XX GT: glucosyltransferase; vaccine; T-cell; B-cell; reaction;

KW Immunoresponse: peptidyl core matrix; dental caries; diptheria;

KW tetanus; measles; polio.

XX

OS Synthetic.

XX

PN W09322341-A.

XX 11-NOV-1993.

PD

XX 30-APR-1993; 93WO-US04094.

PF

XX 01-MAY-1992; 92US-0877295.

PR

XX (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.

XX Smith DJ, Taubman MA;

PI WPI: 1993-368721/46.

XX

DR Synthetic glucosyl-transferase peptide(s) - provoke T- and B-cell

XX reactions in mammals, and are effective against dental caries

PS Claim 3; Page 23; 38pp; English.

The sequences (AAR43694-98) are subsequences from GT, they provoke T- and B-cell responses in mammals. The can be used to produce a vaccine comprising 2 of these sequences attached to a peptidyl core matrix. They are useful in producing T-cell responses to the GT enzyme causing a reduction of colonisation or accumulation of mutants streptococcal strains in mammals. The vaccines can be used in preventing dental caries.

Query Match	77.7%	Score 87	DB 14	Length 22
Best Local Similarity	77.3%	Predg. No. 2.2e-08		
Matches 17	Conservative	2	Mismatches 3	Indels 0
				Gaps 0

```

QY      1  TGARTINGÖLLYFRANGVÖVG  22
        |||:| |' | | | | | | | |
1  tgaqtikgqklyfkangqvkg  22

```

RESULT	2
AAW34159	
ID	AAW34159 standard; peptide; 22 AA

AC	AAW34159;
XX	
DT	18-FEB-1998 (first entry)
XX	

DE GTF antigenic peptide #4.

KM Glucosyltransferase; GTF; enzyme; immunogen; catalytic domain; vaccine  
 KW surface domain; glucan-binding domain; mutans streptococcal strain;  
 KW immune response; glucan-binding activity; dental caries prevention.

05 Streptococcus mutans

PN US5686075-A.

PD 11-NOV-1997

PF 01-MAY-1992; 92US-0877295

PR 30-APR-1993; 93US-0057162

(FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN

Smith DJ, Taubman MA,

DR WPI; 1997-558089/51

PT Immunogenic compositions containing streptococcal

PT to streptococcal glucosyl:transferase for preventing dental caries

PS Claim 2; Column 13; 11pp; English

CC AAMW34156-W34160 represent immunogenic fragments of the Streptococcus  
CC mutans glucosyltransferase (GTF) enzyme. AAMW34157 and AAMW34158 are from  
CC the catalytic domain of GTF, while AAMW4160 is from the GTF surface  
CC domain. AAMW34156 and AAMW34159 are from the glucan-binding domain of GTF.  
CC These sequences can all be used in the immunogenic composition of the  
CC invention. The composition of the invention can alternatively comprise  
CC one or more of these sequences linked to a lysine core matrix (see  
CC AAMW34161-W34165). A composition comprising one of these sequences can be  
CC administered to a mammal to raise an immune response. In a method for  
CC interfering with the enzymatic activity of streptococcal  
CC glucosyltransferase in a mammal, the immune response results in  
CC reduction of the colonisation or accumulation of mutans streptococcal  
CC strains in the mammal. Compositions containing AAMW34156 specifically  
CC interfere with the glucan-binding activity of the streptococcal  
CC glucosyltransferase. The peptides can also be used in vaccines for

CC	preventing dental caries in mammals.
XX	
SQ	Sequence 22 AA;

Query Match	77.7%	Score 87:	DB 18:	Length 22:
Best Local Similarity	77.3%	Pred. No.	2.2e-08:	
Matches 17:	Conservative	2:	Mismatches 3:	Indels 0:
			Gaps	0

```

QY 1 TGARTINGQLLYFRANGVQVK 22
    |||:||| |||:||| |||
DB 1 tgaqtikgqklyfkangqvkg 22

```

RESULT	3
AAV43338	.
ID	AAV43338 standard; peptide; 22 AA

AC	AAV43338;
XX	
DT	26-JAN-2000 (first entry)
XX	

DE GTF antigenic epitope  
xx

KW GTF; antigenic epitope; glucosyltransferase; immunogenic epitope; teeth;

KW immunogenic composition; streptococcal glucan; antibody response  
KW cariogenic bacteria; mutans streptococci; colonisation; caries;  
KW dental caries; immunisation; therapy.

Synthetic.

PN W09952548-A2

PD 21-OCT-1999

PF 09-APR-1999; 99WO-US07828

PR 10-APR-1998; 98US-0081315

PA (LEES/) LEES A.

PA (SMIT/) SMITH D J

PI Lees A, Taubman MA, Smith DJ;

DR WPI; 1999-620289/53

PT Immunogenic compositions for control of dental caries, based on

PT infants -

PS Claim 11; Page 44; 54pp; English

This sequence represents a glucosyltransferase (GTF) immunogenic epitope, and can be used in the immunogenic composition of the invention. The immunogenic composition (A) comprises at least one streptococcal glucan (SG) and at least one component (1) covalently coupled, (in)directly, to SG. (A) elicits antibodies (Ab) to both SG and (1). (A), and related compositions, are used to induce an antibody response to cariogenic bacteria, collectively called "mutans streptococci" (i.e. any of *Streptococcus mutans*, *S. ratius*, *S. ceticus* or *S. sobrinus*), especially to prevent colonisation, to reduce numbers of bacteria or to reduce incidence of caries, in humans or other animals. The composition can therefore be used for the treatment or prevention of caries (particularly dental caries). Abs raised against (A) can be used: (a) to detect SG, and (b) for passive immunisation, e.g. applied directly to the surface of teeth. SG is a poorly immunogenic type 2 T-cell independent antigen but when coupled to (1) it produces a conjugate which induces significant immunity to both components, resulting in immunological memory and long-lasting antibody production against SG, in adults and children.

Sequence 22 AA

Query Match 77.7%: Score 87; DB 20; Length 22;  
 Best Local Similarity 77.3%: Pred. No. 2.2e-08;  
 Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGARTINGQLLYFRANGVQVK 22  
 |||:| | | | | | | | | |  
 Db 1 tgaqtikgqklyfkangqvkv 22

RESULT 4  
 AAW34164  
 ID AAW34164 standard; peptide: 24 AA.  
 AC AAW34164;  
 XX 18-FEB-1998 (first entry)  
 XX GTF antigenic peptide #4 linked to polylysine core.  
 DE  
 XX Glucosyltransferase; GTF; enzyme; immunogen; catalytic domain; vaccine;  
 surface domain; glucan-binding domain; mutans Streptococcal strain;  
 immune response; glucan-binding activity; dental caries prevention.  
 OS Synthetic.  
 OS Streptococcus mutans.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..22  
 FT /label= GTF antigenic peptide #4 (see AAW34159)  
 FT /note= "attached to the dendritic polylysine core via  
 of the alpha-amino group of Lys(23); a second copy  
 of the antigenic 22-mer is linked to Lys(23) via  
 the omega amino group"  
 FT Modified-site 23  
 FT /note= "Lys(23) is linked to one copy of the antigenic  
 peptide through the alpha-amino group, and to a  
 second copy of the peptide (not shown) via the  
 omega amino group"  
 FT Modified-site 24  
 FT /note= "the alpha amino acid group of Lys(24) forms a  
 peptide linkage with the carboxyl amino group of  
 Lys(23); the omega amino group of Lys(24) forms  
 a peptide bond with a second lys residue  
 analogous to Lys(23)"  
 XX US5686075-A.  
 PN 11-NOV-1997.  
 PD 01-MAY-1992; 92US-0877295.  
 PR 30-APR-1993; 93US-0057162.  
 PR 01-MAY-1992; 92US-0877295.  
 XX (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.  
 PA Smith DJ, Taubman MA;  
 PI WPI; 1997-558089/51.  
 DR WPI; 1997-558089/51.  
 XX Immunogenic compositions containing streptococcal  
 PT glucosyltransferase peptide(s) - used for provoking immune response  
 PT to streptococcal glucosyltransferase for preventing dental caries  
 XX Claim 12: Column -; 11pp; English.  
 PS AAW34161-W34165 represent the Streptomyces mutans glucosyltransferase  
 CC (GTF) enzyme immunogenic fragments shown in AAW34156-W34160 linked to a  
 CC polylysine core. AAW34157 and AAW34158 are from the catalytic domain of  
 CC GTF, while AAW34160 is from the GTF surface domain. AAW34156 and AAW34159  
 CC are from the glucan-binding domain of GTF. These sequences, and the  
 CC immunogenic fragments shown in AAW34156-W34160 can all be used in the

CC immunogenic composition of the invention. A composition comprising one of  
 CC these sequences can be administered to a mammal to raise an immune  
 CC response, in a method for interfering with the enzymatic activity of  
 CC streptococcal glucosyltransferase in a mammal. The immune response  
 CC results in reduction of the colonisation or accumulation of mutans  
 CC streptococcal strains in the mammal. Compositions containing AAW34156  
 CC specifically interfere with the glucan-binding activity of the  
 CC streptococcal glucosyltransferase. The peptides can also be used in  
 CC vaccines for preventing dental caries in mammals.  
 XX Sequence 24 AA;  
 SQ

Query Match 77.7%: Score 87; DB 18; Length 24;  
 Best Local Similarity 77.3%: Pred. No. 2.5e-08;  
 Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGARTINGQLLYFRANGVQVK 22  
 |||:| | | | | | | | | |  
 Db 1 tgaqtikgqklyfkangqvkv 22

RESULT 5  
 AAR32925  
 ID AAR32925 standard; Protein: 1592 AA.  
 XX AAR32925;  
 AC 28-JUN-1993 (first entry)  
 DT Glucosyltransferase I.  
 DE GT-1; Streptococcus; dental; caries.  
 XX Streptococcus sobrinus.  
 OS Jp05023188-A.  
 PN 02-FEB-1993.  
 PD 25-JUL-1991; 91JP-0186592.  
 PF 25-JUL-1991; 91JP-0186592.  
 PR 25-JUL-1991; 91JP-0186592.  
 XX (FUKU/) FUKU I.  
 PA (KATO/) KATO K.  
 XX WPI; 1993-079449/10.  
 DR N-PSDB; ANQ37760.  
 DR DNA sequence glucosyltransferase-I - comprises Streptococcus  
 PT sobrinus DNA sequence with at least one nucleotide added or  
 PT deleted  
 XX Claim 13: "acc 15; 29pp; Japanese.  
 PS The DNA sequence from Streptococcus sobrinus strain 6715 encodes  
 CC glucosyltransferase-I (and mutants). The DNA was obtd. by treating  
 CC S. sobrinus 6715 with mutanolysin, extracting the chromosomal DNA,  
 CC partially digesting with Sau3AI and fractionating on agarose gel.  
 CC The 3-5 kbp fragment was ligated into pUC18 and E. coli JM109  
 CC transformed with it. A GT-1 expressing clone was isolated and  
 CC sequenced. The clone may be used in the development of a drug for  
 CC dental caries.  
 XX Sequence 1592 AA;  
 SQ

Query Match 77.7%: Score 87; DB 14; Length 1592;  
 Best Local Similarity 77.3%: Pred. No. 4.5e-06;  
 Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGARTINGQLLYFRANGVQVK 22

Db 1298 tgaqlirgklyfkangqvk 1319

# RESULT 6

AA891047

ID AA891047 standard; Protein: 1577 AA.

AA891047;

22-MAY-1996 (first entry)

Alpha-D-glucosyltransferase.

Alpha-D-glucosyltransferase; primer-independent; soluble glucan;

sucrose; transgenic plant; cloning; Escherichia coli;

phage lambda-C13; vector; plasmid pSG501; plasmid pSG502;

gene transfer; crop improvement; storage carbohydrate; pasture;

feedstuff; senescence; dextran; binder; food; pharmaceutical.

Streptococcus salivarius strain ATCC 25975.

WO9606173-A1.

29-FEB-1996.

24-AUG-1995; 95WO-AU00527.

24-AUG-1994; 94AU-0007643.

(GIFP/) GIFPARD P M.

(JACO/) JACQUES N A.

(SIMP/) SIMPSON C L.

Giffard PM, Jacques NA, Simpson CL;

WPI: 1996-151376/15.

N-PSDB: AAT13139.

Plants contg. new bacterial DNA encoding glucosyl transferase

actively - retain higher levels of stored carbohydrate(s) in a form

readily digestible by ruminants

Claim 4; Page 16-20; 31pp; English.

The sequence represents an alpha-D-glucosyltransferase from

Streptococcus salivarius. The enzyme is primer-independent, and

produces soluble glucan from sucrose. A gene encoding the enzyme

may be cloned and expressed in Escherichia coli using a subclone

of phage lambda-C13, e.g. plasmid pSG501 or plasmid pSG502. The

DNA may also be expressed in a transgenic plant. To improve the

level of stored carbohydrate in a pasture plant which normally

contains low levels, or to prevent degradation of stored carbohydrate

during plant senescence. Dextran may be isolated from the plant, for

use as a food binder or pharmaceutical additive. Primer independence

ensures that the enzyme will be functional in plants. The glucan is

poorly degraded in plants but easily degraded by bacteria in the rumen

of grazing livestock.

Sequence 1577 AA;

Query Match 58.9%; Score 66; DB 17; Length 1577;

Best Local Similarity 54.5%; Pred. No. 0.02;

Matches 12; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGARTINGOLLYFRANGVOYKG 22

Db 1469 tglqnlhmkvyyfsgngaqvk 1490

# RESULT 7

AA810667

ID AA810667 standard; Protein: 2057 AA.

AA810667;

19-JAN-2001 (first entry)

L. mesenteroides alternan sucrose protein.

Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;

syryp.

Leuconostoc mesenteroides.

DE19905069-A1.

10-AUG-2000.

08-FEB-1999; 99DE-1005069.

08-FEB-1999; 99DE-1005069.

(PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.

(PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

Kossmann J, Welsh T, Quanz M, Knuth K;

WPI: 2000-550294/51.

N-PSDB: AAA97904.

New nucleic acid encoding recombinant Leuconostoc mesenteroides

alternan sucrose protein and methods of alternan and fructose

production

Claim 1a; Page 30-36; 64pp; German.

This invention describes a novel nucleic acid molecule (I) encoding an

alternan sucrose (E.C. 2.4.1.140 - an enzyme, that belongs to the

glucosyltransferase group) The recombinant, purified alternan sucrose

gene is useful for the fermentative production of alternan (a

carbohydrate) and/or fructose by secreting the enzyme into a

saccharose-containing culture medium. Alternatively, the enzyme is

contacted with a saccharose-containing solution. The alternan and/or

fructose is then isolated from the medium. Cosmetic products or

foodstuffs containing alternan can be produced. Recombinant production of

alternan sucrose is advantageous as it provides a cost effective means of

producing fructose for high fructose containing syrups, production of

which previously has been achieved by costly production from maize

starch. This sequence represents the Leuconostoc mesenteroides alternan

sucrose protein which is described in the method of the invention.

Sequence 2057 AA;

Query Match 58.0%; Score 65; DB 21; Length 2057;

Best Local Similarity 57.1%; Pred. No. 0.041;

Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 GARTINGOLLYFRANGVOYKG 22

Db 261 glqldnlglyfngqvqvk 281

# RESULT 8

AA87969

ID AA87969 standard; peptide: 320 AA.

AA87969;

29-APR-1999 (first entry)

Toxin A immunogenic fragment p5/6.

Toxin A; immunogenic fragment; vaccine; diagnosis; infection.

XX Clostridium difficile.  
 OS  
 XX  
 PN MO9859053-A1.  
 XX  
 PD 30-DEC-1998.  
 XX  
 PF 19-JUN-1998; 98WO-GB01805.  
 XX  
 PR 07-JAN-1998; 98GB-0000321.  
 PR 20-JUN-1997; 97GB-0013146.  
 XX  
 PA (UNIO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.  
 PA (QUEE-) QUEEN MARY & WESTFIELD COLLEGE.  
 XX  
 PI Douce G, Dougan G, Ward SJ, Wren BW;  
 XX  
 DR WPI; 1999-081281/07.  
 XX  
 PT New isolated Clostridium difficile proteins - comprise fragments of  
 toxin A, used to develop agents for prophylaxis, treatment or  
 diagnosis of C. difficile infections  
 XX  
 PS Claim 1; Fig 6; 82pp; English.  
 XX  
 CC This sequence represents an immunogenic fragment of Clostridium difficile  
 CC (CD) Toxin A, designated fragment p5/6. This sequence is encoded by  
 CC nucleotides 7153-8118 of the Toxin A gene. The invention relates to  
 CC fragments of toxin A of CD which are non-toxic and immunogenic. They can  
 CC be used in the preparation of vaccines against CD infection. The amino  
 CC acid molecules and antibodies against them can be used in the preparation  
 CC of an agent for the prophylaxis or treatment of a CD infection. They can  
 CC also be used for detection and diagnosis of CD infection.  
 CC  
 SQ Sequence 320 AA;  
 XX  
 Query Match 44.2%; Score 49.5; DB 20; Length 320;  
 Best Local Similarity 52.2%; Pred. No. 2;  
 Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;  
 QY 1 TGARTINGQLLYFRAN-GVOYKG 22  
 |||||: || | | | |  
 Db 74 tglrtldgkkyfntntavtvg 96  
 XX  
 RESULT 9  
 AAM87970  
 AAM87970 standard; peptide: 457 AA.  
 AAM87970:  
 XX 29-APR-1999 (first entry)  
 DT  
 XX  
 DE Toxin A immunogenic fragment p5/7.  
 XX  
 KM Toxin A; immunogenic fragment; vaccine; diagnosis; infection.  
 XX  
 OS Clostridium difficile.  
 XX  
 PN MO9859053-A1.  
 XX  
 PD 30-DEC-1998.  
 XX  
 PF 19-JUN-1998; 98WO-GB01805.  
 XX  
 PR 07-JAN-1998; 98GB-0000321.  
 PR 20-JUN-1997; 97GB-0013146.  
 XX  
 PA (UNIO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.  
 PA (QUEE-) QUEEN MARY & WESTFIELD COLLEGE.  
 XX  
 PI Douce G, Dougan G, Ward SJ, Wren BW;

XX WPI; 1999-081281/07.  
 DR  
 XX  
 PT New isolated Clostridium difficile proteins - comprise fragments of  
 PT toxin A, used to develop agents for prophylaxis, treatment or  
 PT diagnosis of C. difficile infections  
 XX  
 PS Example 1; Fig 7; 82pp; English.  
 XX  
 CC This sequence represents an immunogenic fragment of Clostridium difficile  
 CC (CD) Toxin A, designated fragment p5/7. This sequence is encoded by  
 CC nucleotides 6748-8118 of the Toxin A gene. The invention relates to  
 CC fragments of toxin A of CD which are non-toxic and immunogenic. They can  
 CC be used in the preparation of vaccines against CD infection. The amino  
 CC acid molecules and antibodies against them can be used in the preparation  
 CC of an agent for the prophylaxis or treatment of a CD infection. They can  
 CC also be used for detection and diagnosis of CD infection.  
 CC  
 SQ Sequence 457 AA;  
 XX  
 Query Match 44.2%; Score 49.5; DB 20; Length 457;  
 Best Local Similarity 52.2%; Pred. No. 3.1;  
 Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;  
 QY 1 TGARTINGQLLYFRAN-GVOYKG 22  
 |||||: || | | | |  
 Db 211 tglrtldgkkyfntntavtvg 233  
 XX  
 RESULT 10  
 AAR95014  
 AAR95014 standard; Protein: 811 AA.  
 XX  
 AC AAR95014;  
 XX  
 DT 08-JUL-1996 (first entry)  
 XX  
 DE C. difficile toxin A (aa1870-2680).  
 XX  
 KM Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin;  
 KW diarrhoea; diagnosis; therapy.  
 XX  
 OS Clostridium difficile VPI strain 10463 (ATCC 10463).  
 XX  
 XX MO9612802-A1.  
 XX  
 PN  
 PD 02-MAY-1996.  
 XX  
 PF 23-OCT-1995; 95WO-US13737.  
 XX  
 PR 07-JUN-1995; 95US-0480604.  
 PR 24-OCT-1994; 94US-0329154.  
 PR 16-MAR-1995; 95US-0405496.  
 PR 14-APR-1995; 95US-0422711.  
 XX  
 PA (OPHI-) OPHIDIAN PHARM INC.  
 XX  
 PI Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;  
 PI Williams JA;  
 XX  
 DR WPI; 1996-230603/23.  
 XX  
 PT Fusion proteins comprising non-toxin protein and part of toxin  
 PT useful to form anti-toxins against Clostridium botulinum type A, and  
 PT C. difficile type toxins, and to treat C. difficile intoxication,  
 PT partic. diarrhoea  
 XX  
 PS Claim 53; Page 310-312; 434pp; English.  
 XX  
 CC A recombinant protein (AAR95014) comprises amino acids 1870-2680 of  
 CC Clostridium difficile VPI strain 10463 toxin A (see also AAR95016),  
 CC a cytotoxin associated with diarrhoeic disease. It was obt'd. by

CC PCR amplification of genomic DNA, and was expressed as a soluble  
CC fusion protein, with maltose binding protein as fusion partner, in  
CC transformed E. coli host cells. The soluble fusion protein can be  
CC used to raise avian antibodies useful as antitoxins or diagnostics.  
XX  
SQ Sequence 811 AA:

Query Match 44.2%; Score 49.5; DB 17; Length 811;  
Best Local Similarity 52.2%; Pred. No. 6.4;  
Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

OY 1 TGARTINGOLLYFRAN-GVQVKG 22  
|||:|:| | | | |  
Db 587 tglrtldgkkyfntntavavtg 609

RESULT 11

AA95017 standard; Protein; 812 AA.

AA95017;

08-JUL-1996 (first entry)

C. difficile toxin A (aa1873-2684).

Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin; vaccine;  
immunogen; diarrhoea; diagnosis; therapy.

Clostridium difficile VPI strain 10463 (ATCC 10463).

WO9612802-A1.

02-MAY-1996.

23-OCT-1995; 95WO-US13737.

07-JUN-1995; 95US-0480604.

24-OCT-1994; 94US-0329154.

16-MAR-1995; 95US-0405496.

14-APR-1995; 95US-0422711.

(OPHI-) OPHIDIAN PHARM INC.

Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS.

Fusion proteins comprising non-toxin protein and part of toxin  
useful to form anti-toxins against Clostridium botulinum type A, and  
C. difficile type toxins, and to treat C. difficile intoxication,

partic. diarrhoea

Claim 79; Page 354-356; 434pp; English.

A protein (AA95017) comprising amino acids 1873-2684 of Clostridium  
difficile VPI strain 10463 toxin A (see also AA95016) was obtd. by  
PCR amplification of genomic DNA. The PCR product can be expressed  
in transformed E. coli host cells as a soluble fusion protein, with  
maltose binding protein or a polystyridine affinity tag as  
fusion partner. The resulting soluble fusion proteins, which are  
substantially endotoxin-free, can be used in anti-clostridial  
vaccines or to raise avian antibodies useful in clostridial antitoxin  
therapy for humans and animals.

Sequence 812 AA:

Query Match 44.2%; Score 49.5; DB 17; Length 812;  
Best Local Similarity 52.2%; Pred. No. 6.4;  
Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

OY 1 TGARTINGOLLYFRAN-GVQVKG 22  
|||:|:| | | | |  
Db 588 tglrtldgkkyfntntavavtg 610

RESULT 12

AA87971 standard; peptide; 862 AA.

AA87971;

29-APR-1999 (first entry)

Toxin A immunogenic fragment p9/10.

Toxin A; immunogenic fragment; vaccine; diagnosis; infection.

Clostridium difficile.

WO9859053-A1.

30-DEC-1998.

19-JUN-1998; 98WO-GB01805.

07-JAN-1998; 98GB-0000321.

20-JUN-1997; 97GB-0013146.

(UNLO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.

(QUEB-) QUEEN MARY & WESTFIELD COLLEGE.

Douce G, Dougan G, Ward SJ, Wren BW;

WPI; 1999-081281/07.

New isolated Clostridium difficile proteins - comprise fragments of  
toxin A, used to develop agents for prophylaxis, treatment or  
diagnosis of C. difficile infections

Example 1; Fig 8; 82pp; English.

This sequence represents an immunogenic fragment of Clostridium difficile  
(CD) Toxin A, designated fragment p9/10. This sequence is encoded by  
nucleotides 5530-8115 of the Toxin A gene. The invention relates to  
fragments of toxin A of CD which are non-toxic and immunogenic. They can  
be used in the preparation of vaccines against CD infection. The amino  
acid molecules and antibodies against them can be used in the preparation  
of an agent for the prophylaxis or treatment of a CD infection. They can  
also be used for detection and diagnosis of CD infection.

Sequence 862 AA:

Query Match 44.2%; Score 49.5; DB 20; Length 862;  
Best Local Similarity 52.2%; Pred. No. 6.9;  
Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

OY 1 TGARTINGOLLYFRAN-GVQVKG 22  
|||:|:| | | | |  
Db 617 tglrtldgkkyfntntavavtg 639

RESULT 13

AA830543 standard; Protein; 866 AA.

AA830543;

06-MAR-2001 (first entry)

A recombinant protein of toxin A repeating units (RARU).

XX

KM		pathogenic microorganism; nosocomial pathogenic microorganism; vaccine;
KW		immune response.
OS	Clostridium difficile.	
XX		
FH	Key Location/Qualifiers	
FT	Misc-difference 794	/note= "Gln encoded by GGTCAA"
FT	Misc-difference 821..822	/note= "these residues encoded by Gga"
FT	Misc-difference 836..866	/note= "the nucleotides encoding these residues are not given"
PN	WO200061761-A2.	
XX		
PD	19-OCT-2000.	
PF	10-APR-2000; 2000MO-USO9523.	
PR	09-APR-1999; 99US-0128686.	
PR	01-MAR-2000; 2000US-0186201.	
PA	(TECH-) TECHLAB INC.	
PI	(USSH ) US DEPT HEALTH & HUMAN SERVICES.	
PI	Robbins JB;	
DR	WPt; 2000-679490/66.	
PT	Immunogenic compositions useful as vaccines comprise a recombinant protein of toxin A or B of Clostridium difficile conjugated to a polysaccharide of a microbial pathogen	
PS	Example 1; Fig 3; 45pp; English.	
XX		
CC	The present sequence represents a Clostridium difficile recombinant protein comprising toxin A repeating units (rTAU). The protein is used to prepare immunogenic compositions of the invention. The immunogenic compositions comprise a recombinant protein and a polysaccharide component, in which the recombinant protein is encoded by a gene from a strain of Clostridium difficile and the polysaccharide component is isolated from a strain of pathogenic microorganism or is chemically synthesized. The immunogenic compositions are useful for eliciting a protective immune response (T-cell dependent or T-cell independent, a cellular or humoral immune response) to a strain of pathogenic microorganism. The immunogenic composition also elicits a protective immune response against the polysaccharide produced by a strain of a nosocomial pathogenic microorganism. The immunogenic compositions are useful as vaccines for humans, particularly children and animals in affording protection against one or more microbial pathogens.	
SQ	Sequence 866 AA:	
OY	Query Match 44.2%; Score 49.5; DB 21; Length 866; Best Local Similarity 52.2%; Pred. No. 6.9; Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;	
DB	1 TGARTINGOLLYFRAN-GVQYKG 22    ::    617 tglrtldgkkyfntltavtvg 639	
RESULT 15		
ID	AAR95016	
AC	AAR95016 standard; Protein: 2710 AA.	
TX	AAR95016;	
TX	08-JUL-1996 (first entry)	

DE C. difficile toxin A.

XX  
KW Toxin A: cytotoxin; enterotoxin; fusion protein; antitoxin;  
XX vaccine; diarrhoea; diagnosis; therapy.

XX OS Clostridium difficile VPI strain 10463 (ATCC 10463).

XX PN W09612802-A1.

XX PD 02-MAY-1996.

XX PF 23-OCT-1995; 95WO-US13737.

XX PR 07-JUN-1995; 95US-0480604.

PR 24-OCT-1994; 94US-0329154.

PR 16-MAR-1995; 95US-0405496.

PR 14-APR-1995; 95US-0422711.

XX (OPHI-) OPHIDIAN PHARM INC.

PI Firca JR, Klink JA, Padhye NV, Stafford DC, Thalley BS.  
PI Williams JA;

XX WPI; 1996-230603/23.

DR N-PSDB; AAT29248.

XX Fusion proteins comprising non-toxin protein and part of toxin

PT useful to form anti-toxins against Clostridium botulinum type A, and

PT C. difficile type toxins, and to treat C. difficile intoxication,

PT partic. diarrhoea

PS Claim 63; Page 290-302; 434pp; English.

XX Clostridium difficile VPI strain 10463 toxin A (AAR95016), the

CC product of the toxin A gene (AAT29248), is a potent cytotoxin that

CC plays a direct role in damaging gastrointestinal tissues and is

CC associated with diarrhoeic disease. It can be obtd. by expression in

CC transformed E. coli hosts of portions of DNA that together cover the

CC entire toxin A gene. Toxin A, and portions of it (see also

CC AAR95014-15 and AAR95017), pref. expressed as fusions to polystyridine

CC affinity tags or maltose binding protein, are used to raise avian

CC antibodies useful as antitoxins or diagnostics, and in vaccine prodn.

XX Sequence 2710 AA;

XX Query Match 44.2%; Score 49.5; DB 17; Length 2710;

XX st Local Similarity 52.2%; Pred. No. 29;

XX Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

QY 1 TGARTINGQLYFRAN-GYQVKG 22

DB 2460 tglrtldgkkyfntcavavcg 2482

Search completed: March 27, 2002, 13:57:54

Job time: 522 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 14:27:01 ; Search time 53.4 Seconds

(without alignments)  
15.105 Million cell updates/sec

Title: US-09-290-049a-3

Perfect score: 112  
Sequence: 1 TGARTINGOLLYFRANGVOVKG 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	93.8	1476	1	GTFB_STRMU
2	87	77.7	1592	1	GTF2_STRDO
3	87	77.7	1597	1	GTF1_STRDO
4	78	69.6	1375	1	GTF3_STRMU
5	75	67.0	1365	1	GTF5_STRDO
6	67	59.8	1462	1	GTFD_STRMU
7	49.5	44.2	529	1	Y119_LNPOV
8	49.5	44.2	2710	1	TOXA_CLODI
9	47	42.0	126	1	DOC_BPPI
10	47	42.0	788	1	SPL_HUMAN
11	47	42.0	231	1	DLH1_AQUAE
12	46	41.1	231	1	USF_AQOPY
13	46	41.1	245	1	TM24_METJA
14	46	41.1	245	1	LECI_MOUSE
15	46	39.3	1396	1	VLF_BPTS
16	44	38.4	110	1	YC27_ARCEU
17	43	38.4	299	1	Y002_CARBL
18	43	38.4	467	1	ISP6_SCHPO
19	43	38.4	401	1	HAT2_YEAST
20	41.5	37.1	472	1	CBIR_HUMAN
21	41.5	37.1	472	1	CBIR_MOUSE
22	41.5	37.1	473	1	CBIR_MOUSE
23	41.5	37.1	473	1	CBIR_MOUSE
24	41.5	37.1	473	1	CBIR_MOUSE
25	41.5	37.1	473	1	CBIR_MOUSE
26	41.5	37.1	473	1	CBIR_MOUSE
27	41.5	37.1	473	1	CBIR_MOUSE
28	41.5	37.1	473	1	CBIR_MOUSE
29	41.5	37.1	473	1	CBIR_MOUSE
30	41.5	37.1	473	1	CBIR_MOUSE
31	41.5	37.1	473	1	CBIR_MOUSE
32	41.5	37.1	473	1	CBIR_MOUSE
33	41.5	37.1	473	1	CBIR_MOUSE

## ALIGNMENTS

RESULT ID	GTFB_STRMU	STANDARD	PRT: 1476 AA.
AC	P08987: 069381: 069384: 069387: 069390: 069396:		
DT	01-NOV-1988 (Rel. 09, Created)		
DT	15-JUL-1999 (Rel. 38, Last sequence update)		
DT	15-JUL-1999 (Rel. 38, Last annotation update)		
DE	GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)		
DE	(SUCROSE 6-GLUCOSYLTRANSFERASE).		
GN	GTFB.		
OS	Streptococcus mutans.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		
OC	Streptococcus.		
OX	NCBI_TaxID=1309;		
OX	111		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-GS-5.		
RX	MEDLINE=87308013; PubMed=3040685;		
RA	Shiroza T., Ueda S., Kuramitsu H.K.;		
RT	"Sequence analysis of the gltf gene from Streptococcus mutans.";		
RL	J. Bacteriol. 169:4263-4270(1987).		
RN	121		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-MT4239, MT4245, MT4251, MT4467, AND MT9148;		
RX	MEDLINE=98231643; PubMed=9570124;		
RA	Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,		
RA	Kiura S., Hamada S.;		
RT	"Molecular analyses of glucosyltransferase genes among strains of		
RT	Streptococcus mutans.";		
RL	FEBS Microbiol. Lett. 161:331-336(1998).		
CC	-1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT		
CC	TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE		
CC	OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE		
CC	AGGREGATION OF BACTERIAL CELLS AND FOOD DEBR.		
CC	-1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) =		
CC	D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).		
CC	-1- SUBCELLULAR LOCATION: SECRETED.		
CC	-1- DISEASE: DENTAL CARIES.		
CC	-1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA		
CC	1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES). GTF-S SYNTHESIZES		
CC	WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH		
CC	FORMS OF GLUCANS.		
CC	-1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-		
CC	BINDING PROTEIN FROM S. MUTANS.		
CC	-1- This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).		
DR	EMBL; M17361; AAA88588.1; -		P27724 synechocyst
DR	EMBL; D88651; BAA26101.1; -		P32375 saccharomyc
DR	EMBL; D88654; BAA26105.1; -		O23656 crithidia f
			P39814 bacillus su
			P10070 homo sapien
			P56971 poephilla gu
			O59128 pyrococcus
			O50428 methyllobaci
			P47007 saccharomyc
			P32426 streptomyc
			G92184 gallus gall

```

DR EMBL: D88657; BAA26109.1; -
DR EMBL: D88660; BAA26113.1; -
DR EMBL: D89977; BAA26119.1; -
DR PIR: B3135; B3135.
DR InterPro: IPR002479; CW_binding_70.
DR Pfam: PF01473; CW_binding_1; 13.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1
FT CHAIN 34
FT DOMAIN 35 1476
FT DOMAIN 35 1051
FT DOMAIN 1097 1476
FT REPEAT 1097 1130
FT REPEAT 1161 1470
FT REPEAT 1161 1210
FT REPEAT 1225 1275
FT REPEAT 1290 1340
FT REPEAT 1355 1405
FT REPEAT 1420 1470
FT REPEAT 62 62
FT VARIANT 62 62
FT VARIANT 65 65
FT VARIANT 68 68
FT VARIANT 78 78
FT VARIANT 86 86
FT VARIANT 89 89
FT VARIANT 168 168
FT VARIANT 276 276
FT VARIANT 399 399
FT VARIANT 474 474
FT VARIANT 512 512
FT VARIANT 519 519
FT VARIANT 701 701
FT VARIANT 708 708
FT VARIANT 938 938
FT VARIANT 952 957
FT VARIANT 963 964
FT VARIANT 968 970
FT VARIANT 1086 1086
FT VARIANT 1158 1158
FT VARIANT 1163 1163
FT VARIANT 1168 1168
FT VARIANT 1182 1182
FT VARIANT 1234 1234
FT VARIANT 1263 1263
FT VARIANT 1263 1263
FT VARIANT 1264 1264
FT VARIANT 1272 1272
FT VARIANT 1329 1329
FT VARIANT 1394 1394
FT VARIANT 1402 1402
FT VARIANT 1459 1459
FT CONFLICT 570 570
FT CONFLICT 800 817
FT CONFLICT 1310 1310
FT SEQUENCE 1476 AA; 165685 MW; 3479B62B0794D98 CRC64;

Query Match          93.8%; Score 105; DB 1; Length 1476;
Best Local Similarity 95.5%; Pred. No. 7.5e-09;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 1 TGARTINGOLLYFRANGVOYKG 22
Db 1301 TGARTINGOLLYFRANGVOYKG 1332

```

```

RESULT 2
ID GTF2_STRDO STANDARD; PRT; 1592 AA.
AC GTF2_STRDO
DT 01-AUG-1992 (Rel. 23. Created)
DT 01-AUG-1992 (Rel. 23. Last sequence update)
DT 15-DEC-1998 (Rel. 37. Last annotation update)
DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
OC Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxId=1317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6715;
RX MEDLINE=91123227; PubMed=1704006;
RA Abo H., Matsunura T., Kodama T., Ohta H., Fukui K., Kato K.,
RA Kagawa H.;
RT "Peptide sequences for sucrose splitting and glucan binding within
RT streptococcus sobrinus glucosyltransferase (water-insoluble glucan
RT synthetase).";
RL J. Bacteriol. 173:989-996(1991).
CC -I- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS. THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -I- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -
CC -I- D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).
CC -I- SUBCELLULAR LOCATION: SECRETED.
CC -I- DISEASE: DENTAL CARIES.
CC -I- MISCELLANEOUS: GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -I- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as there are no content in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL: D90213; BAA14241.1; -
DR PIR: A38175; A38175.
DR HSP: P00695; 2HRE.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 16.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1
FT CHAIN 38
FT DOMAIN 39 1592
FT DOMAIN 1093 1592
FT DOMAIN 1093 1142
FT REPEAT 1158 1207
FT REPEAT 1222 1272
FT REPEAT 1287 1337
FT REPEAT 1402 1451
FT REPEAT 1514 1563
FT REPEAT 1577 1592
FT SEQUENCE 1592 AA; 176167 MW; BC0A66D079351ECF CRC64;

Query Match          77.7%; Score 87; DB 1; Length 1592;
Best Local Similarity 77.3%; Pred. No. 6.2e-06;

```



FT DOMAIN 35 1050 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1126 1375 GLUCAN-BINDING (APPROXIMATE).  
 FT DOMAIN 1126 1375 2.4 A, 1 C AND 1 AC REPEATS.  
 FT REPEAT 1126 1159 A REPEAT.  
 FT REPEAT 1169 1200 A REPEAT.  
 FT REPEAT 1227 1238 C REPEAT.  
 FT REPEAT 1253 1303 A REPEAT.  
 FT REPEAT 1318 1330 A REPEAT (INCOMPLETE).  
 FT REPEAT 1330 1330 A REPEAT (INCOMPLETE).  
 SQ SEQUENCE 1375 AA: 153022 MW: D4B80CBEE0A0ACE13 CRC64;

Query Match 69.6%; Score 78; DB 1; Length 1375;  
 Best Local Similarity 68.2%; Pred. No. 0.00015;  
 Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 TGARTINGQLYFRANGVOYKG 22  
 1264 TGVTFNGQLYKFPNGVOYAKG 1285

RESULT 5  
 GTF5\_STRDO STANDARD: PRT: 1365 AA.  
 ID GTF5\_STRDO  
 AC P29336;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 01-APR-1993 (Rel. 25, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
 DE (SUCCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTF5.  
 OS Streptococcus downei (Streptococcus sobrinus).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1317;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ME28;  
 RA MEDLINE=90316665; PubMed=2142479;  
 RX Gilmore K.S., Russell R.R., Ferretti J.J.;  
 RT "Analysis of the Streptococcus downei gtfS gene, which specifies a  
 glucosyltransferase that synthesizes soluble glucans.";  
 RL Infect. Immun. 58:2452-2458(1990).  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: SUCCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -  
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -1- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF  
 PRIMER GLUCAN UNLIKE GTF-I.  
 CC -1- DISEASE: DENTAL CARIES.  
 CC -1- MISCELLANEOUS: GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA  
 1,6-GLUCOSE).  
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 BINDING PROTEIN FROM S. MUTANS.  
 CC -----  
 CC This SMSS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M30943; AAA26898.1; -  
 DR PIR: A41483; A41483.  
 DR InterPro: IPR002479; CM\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CM\_binding\_1; 10.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.  
 FT SIGNAL 1 36  
 FT CHAIN 37 1365 OR 37 (POTENTIAL).  
 FT GLUCOSYLTRANSFERASE-S.

FT DOMAIN 37 1050 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1083 1365 GLUCAN-BINDING (APPROXIMATE).  
 FT DOMAIN 1083 1365 4.5 X TANDEM REPEATS.  
 FT REPEAT 1083 1131 1.  
 FT REPEAT 1150 1199 2.  
 FT REPEAT 1225 1274 3.  
 FT REPEAT 1289 1339 4.  
 FT REPEAT 1353 1365 5 (INCOMPLETE).  
 SQ SEQUENCE 1365 AA: 151590 MW: 167296B5A2E8C476 CRC64;

Query Match 67.0%; Score 75; DB 1; Length 1365;  
 Best Local Similarity 63.6%; Pred. No. 0.00044;  
 Matches 14; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGARTINGQLYFRANGVOYKG 22  
 DB 1235 TGEOTIDGCKVFFDNGVOYKG 1256

RESULT 6  
 GTFD\_STRMU STANDARD: PRT: 1462 AA.  
 ID GTFD\_STRMU  
 AC P49331; O69383; O69386; O69389; O69392; O69398;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
 DE (SUCCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTFD.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GS-5;  
 RA MEDLINE=9110958; PubMed=2148600;  
 RX Honda O., Kato C., Kuramitsu H.K.;  
 RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding  
 the glucosyltransferase-S enzyme.";  
 RL J. Gen. Microbiol. 136:2099-2105(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MT429; MT4251; MT4467; AND MT8148;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,  
 RA Kimura S., Hamada S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans.";  
 RL FEMS Microbiol. Lett. 161:331-336(1998).  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: SUCCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -  
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- DISEASE: DENTAL CARIES.  
 CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
 WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
 FORMS OF GLUCANS.  
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 BINDING PROTEIN FROM S. MUTANS.  
 CC -----  
 CC This SMSS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

```

CC -----
DR EMBL: M29296; AAA26895.1; -
DR EMBL: D88653; BAA26103.1; -
DR EMBL: D88656; BAA26107.1; -
DR EMBL: D88659; BAA26111.1; -
DR EMBL: D88662; BAA26115.1; -
DR EMBL: D89979; BAA26121.1; -
DR InterPro: IPR002479; CM_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; Glyco_hydro_70; 1.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 1462 2
FT CHAIN ? 1423 1423
FT REPEAT 1232 1295
FT REPEAT 1296 1359
FT REPEAT 1360 1423
FT VARIANT 58 58
FT VARIANT 68 68
FT VARIANT 81 81
FT VARIANT 113 113
FT VARIANT 122 122
FT VARIANT 132 132
FT VARIANT 135 135
FT VARIANT 202 202
FT VARIANT 255 255
FT VARIANT 275 275
FT VARIANT 288 288
FT VARIANT 301 301
FT VARIANT 313 313
FT VARIANT 317 317
FT VARIANT 328 328
FT VARIANT 350 350
FT VARIANT 628 628
FT VARIANT 688 688
FT VARIANT 726 732
FT VARIANT 726 730
FT VARIANT 964 964
FT VARIANT 1019 1019
FT VARIANT 1059 1060
FT VARIANT 1060 1060
FT VARIANT 1080 1080
FT VARIANT 1142 1142
FT VARIANT 1198 1198
FT VARIANT 1220 1220
FT VARIANT 1280 1280
FT VARIANT 1282 1282
FT VARIANT 1290 1290
FT VARIANT 1311 1311
FT VARIANT 1403 1403
FT VARIANT 1425 1425
FT VARIANT 1449 1449
FT CONFLICT 1428 1462
FT SEQUENCE 1462 AA; 163512 MW; 5C6541F0DCB0DF00 CRC64;

Query Match
Best Local Similarity 59.8%; Score 67; DB 1; Length 1462;
Matches 13; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
1 TGARTINGOLLYFRANGOVKG 22
|||||:|||||

```

```

Db 1332 TGSCITACKKLYFASDGKOVKG 1353
RESULT 7
ID Y119_NPVOP STANDARD; PRT; 529 AA.
AC 010358;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 59.0 KDA PROTEIN PRECURSOR (ORF119).
OS Orygia pseudotsugata multicausid polyhedrosis virus (OPMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
CC Nucleopolyhedrovirus.
CX NCBI_TaxID=164623;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA Rohmann G.F.;
RT "The sequence of the Orygia pseudotsugata multicausid polyhedrosis virus genome."
RT Virology 229:381-399(1997).
CC CC
CC -1 SIMILARITY: TO CORRESPONDING ORF IN ACNMPV.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U75930; AAC59118.1; -
KW Hypothetical protein; Signal.
FT SIGNAL 1 20
FT CHAIN 21 529
FT SEQUENCE 529 AA; 58983 MW; E85F81DD9219B0B CMC64;

Query Match
Best Local Similarity 44.2%; Score 49.5; DB 1; Length 529;
Matches 11; Conservative 5; Mismatches 5; Indels 1; Gaps 1;
1 TGARTINGOLLYFRANGOVKG 22
|||||:|||||
Db 260 SGRRT-SGRFVHADGVEVSG 280

RESULT 8
ID TOXA_CLODI STANDARD; PRT; 2710 AA.
AC P16154;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TOXIN A.
GN TOXA OR TCDA.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
CC Clostridium.
CX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VPI 10463;
RX MEDLINE=90221894; PubMed=2109310;
RA Sauerborn M., von Eichel-Streiber C.;
RT "Nucleotide sequence of Clostridium difficile toxin A."
RL Nucleic Acids Res. 18:1629-1630(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=VPI 10463;
RX MEDLINE=90129305; PubMed=2105276;

```

RA Dove C.H., Wang S.Z., Price S.B., Phelps C.J., Lively D.M.,  
RA Wilkins T.W., Johnson J.L.:  
RT "Molecular characterization of the Clostridium difficile toxin A  
RL gene";  
RM Infect. Immun. 58:480-488(1990).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-VPI 10463:  
RA von Elbehl-Streiber C.:  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA  
CC REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE  
CC DIFFERENT OLIGOPEPTIDES.  
CC -1- DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN  
CC ENTEROTOXIN CALLED A AND CYTOTOXIN B. ONLY AFTER THE ENTERAL  
CC DELIVERY OF THE ENTEROTOXIN A MAY THE CHARACTERISTIC DISEASE  
CC CALLED PSEUDOMEMBRANOUS COLITIS BE INDUCED.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: X51797; CAA36094.1; -  
DR EMBL: M30307; AAA23283.1; -  
DR EMBL: X92982; CAA63564.1; -  
DR PIR: S08638; S08638.  
DR InterPro: IPR002479; CW\_binding.  
DR Pfam: PF01473; CW\_binding\_1; 31.  
DR Toxin: Enterotoxin.  
KM SEQUENCE 2710 AA; 308052 MW; 0A6E52CE84C14421 CRC64;

Query Match 44.2%; Score 49.5; DB 1; Length 2710;  
Best Local Similarity 52.2%; Pred. No. 11;  
Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

QY 1 TGARTINGQLYFRAN-GVQVKG 22  
DB 2460 TGLRTIDGRKRYFNTAVAVTG 2482

RESULT 9  
BPPI DOC\_BPPI STANDARD: PRT: 126 AA.  
AC 006259;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE DEATH ON CURING PROTEIN.  
DOC.  
GN Bacteriophage P1.  
OS Viruses: dsDNA viruses, no RNA stage; Tailed phages; Myoviridae.  
OC NCB1\_Taxid-10678;  
OX [1]  
RN SEQUENCE FROM N.A.  
RA MEDLINE=94016561; PubMed=8411153;  
RA Lehnher H., Maguin E., Jafri S., Yarmolinsky M.B.:  
RT "Plasmid addition genes of bacteriophage P1: doc, which causes cell  
RT death on curing of prophage, and phd, which prevents host death when  
RT prophage is retained";  
RL J. Mol. Biol. 233:414-428(1993).  
CC -1- FUNCTION: KILLS CELLS. DOC AND PHD PROTEIN FUNCTION IN UNISSON TO  
CC STABILIZE PLASMID NUMBER BY INDUCING A LETHAL RESPONSE TO PLASMID  
CC LOSS.  
CC -1- MISCELLANEOUS: THE CONCENTRATION OF PHD IN P1 LYSOGENS IS FAR  
CC GREATER THAN THAT OF THE POISON IT ANTAGONIZES, SUCH AN EXCESS MAY  
CC ASSUME THE WELL-BEING OF CARRIERS OF THE ADDICTING PLASMID.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: M95666; AAA16931.1; -  
DR PIR: S38555; S38555.  
SQ SEQUENCE 126 AA; 1358 MW; E04A397538A595CD CRC64;

Query Match 42.0%; Score 47; DB 1; Length 126;  
Best Local Similarity 60.0%; Pred. No. 1.1;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 6 INGQLLYFRANGVQV 20  
DB 77 LNSALLFLRRNGVQV 91

RESULT 10  
SPL\_HUMAN ID SPL\_HUMAN STANDARD: PRT: 785 AA.  
AC P08047; Q9NYE7; Q9H305;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE TRANSCRIPTION FACTOR SPL.  
GN SPL OR TSEPL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCB1\_Taxid-9606;  
RN [1]  
RP SEQUENCE OF 4-785 FROM N.A.  
RC TISSUE=Cervical carcinoma;  
RA Haggart M.H., Ladurner A.G.;  
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 1-558 FROM N.A.  
RA MEDLINE=20545561; PubMed=10973950;  
RA Takahara T., Kanazu S., Yanagisawa S., Akanuma H.;  
RT "Heterogeneous Spl mRNAs in human HepG2 cells include a product of  
RT homotypic trans-splicing";  
RL J. Biol. Chem. 275:38067-38072(2000).  
RN [3]  
RP SEQUENCE OF 90-785 FROM N.A., AND SEQUENCE OF 359-375 AND 670-675.  
RA MEDLINE=88080466; PubMed=3319186;  
RA Kadosaga J.T., Carner K.R., Maslarsz F.R., Tjian R.;  
RT "Isolation of cDNA encoding transcription factor Spl and functional  
RT analysis of the DNA binding domain";  
RL Cell 51:1079-1090(1987).  
RN [4]  
RP O-GLYCOSYLATION.  
RA MEDLINE=89003041; PubMed=3139301;  
RA Jackson S.P., Tjian R.;  
RT "O-glycosylation of eukaryotic transcription factors: implications  
RT for mechanisms of transcriptional regulation";  
RL Cell 55:125-133(1988).  
RN [5]  
RP STRUCTURE BY NMR OF 654-684 AND 684-712.  
RA MEDLINE=97218212; PubMed=9065444;  
RA Narayan V.A., Kiriacki R.W., Cardona J.P.;  
RT "Structures of zinc finger domains from transcription factor Spl.  
RT Insights into sequence-specific protein-DNA recognition";  
RL J. Biol. Chem. 272:7801-7809(1997).  
RN [6]  
RP IDENTIFICATION OF SEROTONIN 1A RECEPTOR PROMOTER BINDING SITES.  
RA MEDLINE=96224025; PubMed=8626793;  
RA Parks C.L., Shenk T.;  
RT "The serotonin 1a receptor gene contains a TATA-less promoter that  
RT responds to MAZ and Spl.";

```

RL J. Biol. Chem. 271:4417-4430(1996).
CC -I- FUNCTION: BINDS TO GC BOX PROMOTERS ELEMENTS AND SELECTIVELY
CC ACTIVATES MRNA SYNTHESIS FROM GENES THAT CONTAIN FUNCTIONAL
CC RECOGNITION SITES. CAN INTERACT WITH G/C-RICH MOTIFS FROM
CC SEROTONIN RECEPTOR PROMOTER.
CC -I- SUBCELLULAR LOCATION: NUCLEAR.
CC -I- PPM: O-GLYCOSYLATED; CONTAINS N-ACETYLGLUCOSAMINE SIDE CHAINS.
CC -I- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/annouce/
CC -----
CC EMBL; AF252284; AAF67726.1; -
CC EMBL; AB039286; BAB13476.1; -
CC EMBL; J03133; AAA61154.1; -
CC PIR; A29635; A29635.
CC PDB; 1SP1; 2I-APR-97.
CC PDB; 1SP2; 2I-APR-97.
CC TRANSFAC; T00759; -
CC GLYCOsiteDB; P08047; -
CC MIM; 189906; -
CC InterPro; IPR000822; Znf-C2H2.
CC DR Pfam; PF00096; Zf-C2H2; 3.
CC DR SMART; SM00355; Znf_C2H2; 3.
CC DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
CC DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
CC Transcription regulation; Activator; Zinc-finger; Metal-binding;
CC DN-binding; Nuclear protein; Repeat; Glycoprotein; 3D-structure.
CC DOMAIN 626 708 ZINC FINGERS.
CC FT ZN_FING 626 650 C2H2-TYPE.
CC FT ZN_FING 656 680 C2H2-TYPE.
CC FT ZN_FING 686 708 C2H2-TYPE.
CC FT CONFLICT 366 366 D -> G (IN REF. 3; AA SEQUENCE).
CC FT CONFLICT 670 670 S -> F (IN REF. 3; AA SEQUENCE).
CC FT SEQUENCE 785 AA; 80693 MW; 43893DBF6518B9EA CRC64;
OY 4 RTINGOLLYFRANGVOVK 21
OY ::::| | | | | | | |
OY 180 QTVDSGQQLQAPATGAQVQ 197

RESULT 11
SPL_RAT STANDARD: PRT; 788 AA.
AC Q01714;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRANSCRIPTION FACTOR SP1.
GN SP1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93010958; PubMed=1356762;
RA Imetaka H., Sogawa K., Yasumoto K., Kikuchi Y., Sasano K.,
RA Kobayashi A., Hayami M., Fujii-Kuriyama Y.;
RA "Two regulatory proteins that bind to the basic transcription element
RA (BRE), a GC box sequence in the promoter region of the rat p-4501A1

```

```

RT  gene ". ;
RL  ENBO J. 11:3663-3671(1992)
CC  -1- FUNCTION: BINDS TO GC BOX PROMOTERS ELEMENTS AND SELECTIVELY
CC  ACTIVATES MNNA SYNTHESIS FROM GENES THAT CONTAIN FUNCTIONAL
CC  RECOGNITION SITES.
CC  -1- SUBCELLULAR LOCATION: NUCLEAR.
CC  -1- PMM: O-GLYCOSYLATED; CONTAINS N-ACETYLGLUCOSAMINE SIDE CHAINS
CC  (BY SIMILARITY).
CC  -1- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
CC  PROTEINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; D12768; BAA0235.1; -
DR  PIR; JS0747; JS0747.
DR  HSSP; P08047; ISP1.
DR  TRANSFAC; T00754; -
DR  InterPro; IPR000822; znf-C2H2.
DR  Pfam; PF00096; zf-C2H2_3.
DR  PRINTS; PR00048; ZINC-FINGER.
DR  SMART; SM00355; znf-C2H2; 3.
DR  PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR  PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
KW  Transcription regulation; Activator; zinc-finger; Metal-binding;
KW  DNA-binding; Nuclear protein; Repeat; Glycoprotein.
FT  DOMAIN 629 711
FT  ZN_FING 629 653 C2H2-TYPE.
FT  ZN_FING 659 683 C2H2-TYPE.
FT  ZN_FING 689 711 C2H2-TYPE.
SQ  SEQUENCE 788 AA; 8105 MW; AA2B0CAB81AB80C CRC64;
.

Query Match 42.0%; Score 47; DB 1; Length 788;
Best Local Similarity 50.0%; Pred. No. 7.7;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 4 RTINGQLLYFRANGYQVR 21
Db 183 QTVDDQLQLOFATGAQVQ 200

RESULT 12
DLHH_AQUAE STANDARD; PRT; 231 AA.
ID DLHH_AQUAE
AC 067802;
DT 15-DEC-1998 (Rel. 37, Created)
PT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PURVATIVE CARBOXYMETHYLENEBUTENOLIDASE (EC 3.1.1.45) (DIENEIACETONE
DE HYDROLASE) (Dlh).
DE AQ_1997.
GN Aquifex aeolicus.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RX MEDLINE=98196666; PubMed=9537320;
RA Decker G., Warren P.V., Gaasterland T., Young M.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus." ;
RL Nature 392:353-358(1998).
CC -1- CATALYTIC ACTIVITY: 4-CARBOXYMETHYLENEBUT-2-EN-4-OLIDE + H(2)O =
CC 4 OXOHED-2-ENEDIOLATE.
CC -1- SIMILARITY: BELONGS TO THE DIENEIACETONE HYDROLASE FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE000767; AAC07773.1; -.
DR InterPro; IPR002925; DLH.
DR InterPro; IPR000379; Est_lip_thioest_actsite.
DR Pfam; PF01738; DLH; 1.
KW Hypothetical protein; Hydrolase; Complete proteome.
FT ACT_SITE 118 118 BY SIMILARITY.
FT ACT_SITE 167 167 BY SIMILARITY.
FT ACT_SITE 199 199 BY SIMILARITY.
FT SEQUENCE 231 AA; 26357 MW; 8268DA32A6980751 CRC64;

```

```

Query Match 41.1%; Score 46; DB 1; Length 231;
Best Local Similarity 53.3%; Pred. No. 3.1;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 8 GOLLYFRANGOVKG 22
   ||:| |||:|
Db 2 GOMVEFEKNGVKRG 16

```

```

RESULT 13
ID USF_A00PY STANDARD: PRT; 231 AA.
AC P46209;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE USF PROTEIN.
GN USF.
OS Aquifex pyrophilus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=2714;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KO15A;
RX MEDLINE=96062250; PubMed=7592443;
RA Behamer W., Shao Z., Mages W., Rachel R., Stetter K.O.,
RA Schmitt R.;
RA "Flagellar structure and hyperthermophily: analysis of a single
RA flagellin gene and its product in Aquifex pyrophilus.";
RL J. Bacteriol. 177:6630-6637(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U17575; AAA8922.1; -.
DR InterPro; IPR002925; DLH.
DR InterPro; IPR000379; Est_lip_thioest_actsite.
DR Pfam; PF01738; DLH; 1.
FT SEQUENCE 231 AA; 26140 MW; C55E40839920195 CRC64;

```

```

Query Match 41.1%; Score 46; DB 1; Length 231;
Best Local Similarity 53.3%; Pred. No. 3.1;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 8 GOLLYFRANGOVKG 22
   ||:| |||:|
Db 2 GOMVEFEKNGVKRG 16

```

```

RESULT 14
ID T2M4_METJA STANDARD: PRT; 245 AA.
AC T2M4_METJA
ID O58723;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TYPE II RESTRICTION ENZYME MJAIIV (EC 3.1.21.4) (ENDONUCLEASE MJAIIV)
DE (R. MJAIIV).
GN MJAIIV OR MJ1327.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Claydon R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Feldman J.F., Fushman J.L., Nguyen D.,
RA Ullrichback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA jannaschii.";
RT Science 273:1058-1073(1996).
RN [2]
RP CHARACTERIZATION.
RA Noren C.J., Roberts R.J., Patel J., Byrd D.R., Morgan R.D.;
RA "Method for screening restriction endonucleases.";
RT Patent number WO9911821, 11-MAR-1999.
CC -I- CATALYTIC ACTIVITY: RECOGNIZES THE DOUBLE-STRANDED UNMETHYLATED
CC SEQUENCE GTNANC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U67573; AAB99343.1; -.
DR REBASE; 3894; MjaIV.
DR TIGR; MJ1327; -.
KW Hydrolyase; Endonuclease; Nuclease; Restriction system;
KW Complete proteome.
FT SEQUENCE 245 AA; 28779 MW; 4597C2B8EF1FA973 CRC64;

```

```

Query Match 41.1%; Score 46; DB 1; Length 245;
Best Local Similarity 45.0%; Pred. No. 3.3;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 1 TGARTINGOLLYFRANGOV 20
   | |||:| | | |
Db 129 TSEITINGELVNRISGNEI 148

```

```

RESULT 15
LECI_MOUSE
ID LECI_MOUSE STANDARD: PRT; 301 AA.
AC P24721;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ASIALOGLYCOPROTEIN RECEPTOR 2 (HEPATIC LECTIN 2) (MHL-2) (ASGP-R)
DE (ASGPR).

```



```

GN ASGR2 OR ASGR-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Liver;
RX MEDLINE=91027942; PubMed=2223888;
RA Sanford J.P., Doyle D.;
RT "Mouse asialoglycoprotein receptor cDNA sequence: conservation of
RL receptor genes during mammalian evolution.";
RL Biochim. Biophys. Acta 1087:259-261(1990).
CC -I- FUNCTION: MEDIATES THE ENDOCYTOSIS OF PLASMA GLYCOPROTEINS TO
CC WHICH THE TERMINAL SIALIC ACID RESIDUE ON THEIR COMPLEX
CC CARBOHYDRATE MOIETIES HAS BEEN REMOVED. THE RECEPTOR RECOGNIZES
CC TERMINAL GALACTOSE AND N-ACETYLGALACTOSAMINE UNITS. AFTER LIGAND
CC BINDING TO THE RECEPTOR, THE RESULTING COMPLEX IS INTERNALIZED AND
CC TRANSPORTED TO A SORTING ORGANELLE, WHERE RECEPTOR AND LIGAND ARE
CC DISASSOCIATED. THE RECEPTOR THEN RETURNS TO THE CELL MEMBRANE
CC SURFACE.
CC -I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -I- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN HEPATIC PARENCHYMAL
CC CELLS.
CC -I- MISCELLANEOUS: CALCIUM IS REQUIRED FOR LIGAND BINDING.
CC -I- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53042; CA37211.1; -.
DR PIR; S13165; S13165.
DR HSSP; P06734; 1KJE.
DR MCD; MGI:88082; Asgr2.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
KW Lectin; Glycoprotein; Receptor; Endocytosis; Transmembrane;
KW Calcium; Signal-anchor; Phosphorylation.
FT DOMAIN 1 58 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 59 79 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 80 301 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 169 295 C-TYPE LECTIN (LONG FORM).
FT DISULFID 170 181 BY SIMILARITY.
FT DISULFID 198 293 BY SIMILARITY.
FT DISULFID 271 285 BY SIMILARITY.
FT CARBOHYD 97 97 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 301 AA; 34907 MW; 3A29F1AFBA68F298 CRC64;

Query Match 41.1%; Score 46; DB 1; Length 301;
Best Local Similarity 50.0%; Pred. No. 4.1;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 4 RTINGOLLYFRANGVO 19
Db 153 RTIFCOLAYFQSNQTE 168

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using .sw model

Run on: March 27, 2002, 14:26:02 ; Search time 188.53 Seconds  
(without alignments)  
17.069 Million cell updates/sec

Title: US-09-290-049a-3

Perfect score: 112

Sequence: 1 TGAATNGOLLYFRANGVQYKG 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 14627329 residues

al number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: SPREMBL\_17:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organelle:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_rhodent:\*  
13: sp\_virus:\*  
14: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	90	80.4	1455	2	069382 streptococc
2	90	80.4	1455	2	069388 streptococc
3	90	80.4	1455	2	069397 streptococc
4	87	77.7	1590	2	059983 streptococc
5	86	76.8	1455	2	069391 streptococc
6	82	73.2	1599	2	000599 streptococc
7	80	71.4	1518	2	000600 streptococc
8	78	69.6	1390	2	069385 streptococc
9	76	67.9	1590	2	055263 streptococc
10	75	67.0	1290	2	048756 streptococc
11	75	67.0	1338	2	09WXA34 streptococc
12	73	65.2	1449	2	055264 streptococc
13	72	64.3	1477	2	09L466 leuconostoc
14	72	64.3	1508	2	052224 leuconostoc
15	72	64.3	1508	2	09EZH5 leuconostoc
16	72	64.3	1577	2	054178 streptococc
17	68	60.7	1449	2	068542 streptococc
18	68	60.7	1449	2	068542 streptococc
19	66	58.9	1577	2	055265 streptococc

20	65.5	58.5	330	2	055228 streptococc
21	65	58.0	2057	2	09PER05 leuconostoc
22	64	57.1	1512	2	09WXA35 streptococc
23	57	50.9	1527	2	09ZAR4 leuconostoc
24	57	50.9	1575	2	09LCH3 streptococc
25	54	48.2	221	5	044353 caenorhabdi
26	50	44.6	831	2	050076 clostridium
27	47	42.0	126	2	047171 escherichia
28	47	42.0	126	2	047172 escherichia
29	47	42.0	126	2	047173 escherichia
30	47	42.0	255	2	066378 clostridium
31	47	42.0	396	2	09EP932 clostridium
32	47	42.0	781	11	089090 mus musculu
33	47	42.0	784	11	089087 mus musculu
34	46	41.1	619	2	054972 streptococc
35	46	41.1	1501	2	006307 alieylobacter
36	46	41.1	2364	2	046342 caulobacter
37	45	40.2	142	2	09ABT5 mus musculu
38	45	40.2	181	11	09C021 streptomyce
39	45	40.2	249	2	09RK49 arabidopsis
40	45	40.2	410	2	032351 arabidopsis
41	45	40.2	587	10	09LDB8 arabidopsis
42	45	40.2	630	10	049501 arabidopsis
43	45	40.2	649	2	003706 clostridium
44	44	39.3	181	4	09ULC4 homo sapien
45	44	39.3	181	11	09DB27 mus musculu

## ALIGNMENTS

RESULT 1	ID	069382	PRELIMINARY;	PRT; 1455 AA.
AC	069382;			
DT	01-AUG-1998 (TREMBLrel. 07, Created)			
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	GLUCOSYLTRANSFERASE-ST.			
GN	GTFC.			
OS	Streptococcus mutans.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_TaxID=1309;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MT8148;			
RX	MEDLINE=98231643; PubMed=9570124;			
RA	Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,			
RA	Kimura S., Hamada S.;			
RT	"Molecular analyses of glucosyltransferase genes among strains of			
RT	Streptococcus mutans.";			
RL	FEMS Microbiol. Lett. 161:331-336(1998).			
DR	EMBL; D88652; BAA26102.1; -			
DR	InterPro: IPR002479; CW_binding.			
DR	InterPro: IPR003318; Glyco_hydro_70.			
DR	PFam: PF01473; CW_binding_1; 9.			
KW	PFam: PF02324; Glyco_hydro_70; 1.			
SO	SEQUENCE 1455 AA; 162969 MW; 27D4D3A1ECCA2939 CRC64;			
QY	1 TGAATNGOLLYFRANGVQYKG 22			
Db	1329 TGAATNGOLLYFRANGVQYKG 1350			
Query Match	80.4%;	Score 90;	DB 2;	Length 1455;
Best Local Similarity	72.7%;	Pred. No. 9.2e-06;		
Matches 16;	Conservative 3;	Mismatches 3;	Indels 0;	Gaps 0;
RESULT 2				
069388				

ID 069388 PRELIMINARY; PRT; 1455 AA.  
 AC 069388;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-ST.  
 GN GTF.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT4239;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
 Kimura S., Hamada S.;  
 PT "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans.";  
 DR EMBL; D88658; BAA26110.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR Pfam: PF01473; CW\_binding\_1; 10.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 SO SEQUENCE 1455 AA; 163045 MW; 6D90A4978D35DD82 CRC64;

Query Match 80.4%; Score 90; DB 2; Length 1455;  
 Best Local Similarity 72.7%; Pred. No. 9.2e-06;  
 Matches 16; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 TGARTINGOLLYFRANGVOYKG 22  
 Db 1329 TGARVNGORLYFRKSNQVAKG 1350

RESULT 3  
 069397 PRELIMINARY; PRT; 1455 AA.  
 AC 069397;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-ST.  
 GN GTF.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT4467;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
 Kimura S., Hamada S.;  
 PT "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans.";  
 DR EMBL; D89978; BAA26120.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR Pfam: PF01473; CW\_binding\_1; 9.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 SO SEQUENCE 1455 AA; 162913 MW; A1263427BF24E8E1 CRC64;

Query Match 80.4%; Score 90; DB 2; Length 1455;  
 Best Local Similarity 72.7%; Pred. No. 9.2e-06;  
 Matches 16; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGARTINGOLLYFRANGVOYKG 22  
 Db 1329 TGARVNGORLYFRKSNQVAKG 1350  
 RESULT 4  
 ID 059983 PRELIMINARY; PRT; 1590 AA.  
 AC 059983;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE  
 6-GLUCOSYLTRANSFERASE).  
 GN GTFI.  
 OS Streptococcus sobrinus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1310;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OM2176;  
 RX MEDLINE=94146405; PubMed=8312602;  
 RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.,  
 RT "DNA sequence of the glucosyltransferase gene of serotype d  
 Streptococcus sobrinus.";  
 RL DNA Seq. 4:19-27(1993).  
 CC -I- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) - D-  
 FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).  
 DR EMBL; D13858; BAA02976.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR Pfam: PF01473; CW\_binding\_1; 16.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Signal; Transferase; Glucosyltransferase.  
 FT SIGNAL 1 38 POTENTIAL.  
 FT CHAIN 39 1590  
 SO SEQUENCE 1590 AA; 175955 MW; C3C83A57CF3C2B0E CRC64;

Query Match 77.7%; Score 87; DB 2; Length 1590;  
 Best Local Similarity 77.3%; Pred. No. 3.1e-05;  
 Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 TGARTINGOLLYFRANGVOYKG 22  
 Db 1296 TGAOTINGOKLYFRKANGVOYKG 1317

RESULT 5  
 069391 PRELIMINARY; PRT; 1455 AA.  
 AC 069391;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-ST.  
 GN GTF.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT4251;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
 Kimura S., Hamada S.;  
 PT "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans.";  
 DR EMBL; D88661; BAA26114.1; -  
 DR InterPro: IPR002479; CW\_binding.

Query Match	73.28;	Score 82;	DB 2;	Length 1599;
Best Local Similarity	77.38;	Pred. No. 0.0002;		

GN	Streptococcus mutans
OS	Streptococcus mutans

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT4245;  
RX MEDLINE-98231643; PubMed-9570124;  
RA Fujiwara T., Teruo Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
Kimura S., Hamada S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
RT Streptococcus mutans.";  
RL EMBL, Microbiol. Lett. 161:331-336(1998).  
DR EMBL, D88655; BAA26106.1;  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 7.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SEQUENCE 1390 AA; 155375 MW; 8847E4956EF05E9F CRC64;

Query Match 69.6%; Score 78; DB 2; Length 1390;  
Best Local Similarity 68.2%; Pred. No. 0.0075;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 TGARTINGOLTYFRANGVOVGK 22  
DB 1264 TGTVTNGCRLXFKRNGVAKG 1285

RESULT 9  
O55263  
ID O55263; PRELIMINARY; PRT: 1590 AA.  
AC O55263;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE GTF-I.  
OS Streptococcus sobrinus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1310;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 33478;  
RA Sato S.;  
Ann. Kagoshima Univ. Dental School 16:23-29(1996).  
EMBL, D63570; BAA09792.1;  
InterPro: IPR002479; CW\_binding.  
InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 15.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW Transferase  
SEQUENCE 1590 AA; 176057 MW; 9DF7A3F2C6E4FD43 CRC64;

Query Match 67.9%; Score 76; DB 2; Length 1590;  
Best Local Similarity 68.2%; Pred. No. 0.0018;  
Matches 15; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 TGARTINGOLTYFRANGVOVGK 22  
DB 1231 TGAQTGKOKLYFANGGEQVGK 1252

RESULT 10  
O48756  
ID O48756; PRELIMINARY; PRT: 1290 AA.  
AC O48756;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE DEXTRANSUCRASE.

OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NRRL B1299;  
RX MEDLINE-97136686; PubMed-8982063;  
RA Monchois V., Willemot R.M., Renaud-Simeon M., Croux C., Monsan P.;  
RT "Cloning and sequencing of a gene coding for a novel dextransucrase  
RT from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-  
RT 6) and alpha (1-3) linkages.";  
RL Gene 182:23-32(1996).  
DR EMBL, U38181; AAB40875.1;  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 11.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW  
SEQUENCE 1290 AA; 145590 MW; 355CCE96B749FMA CRC64;

Query Match 67.0%; Score 75; DB 2; Length 1290;  
Best Local Similarity 68.2%; Pred. No. 0.0021;  
Matches 15; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 TGARTINGOLTYFRANGVOVGK 22  
DB 1245 TGYVINGQTLTFDADGROYKG 1266

RESULT 11  
O9WXJ4  
ID O9WXJ4; PRELIMINARY; PRT: 1338 AA.  
AC O9WXJ4;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE GTF-S.  
GN GTF-S.  
OS Streptococcus criceti.  
OG Plasmid pAM1.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1333;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-HS-6;  
RA Inoue M., Fukui K., Miyagi A.;  
RT "S.cricetus glucosyltransferase(gtfs and gtfS) genes.";  
RT Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL, AB026123; BAA77236.1;  
DR HSP, P06278; 1VJ5.  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 10.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW Plasmid.  
SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;

Query Match 67.0%; Score 75; DB 2; Length 1338;  
Best Local Similarity 63.6%; Pred. No. 0.0022;  
Matches 14; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGARTINGOLTYFRANGVOVGK 22  
DB 1207 TGSQTAGQKVEFPNGVOVGK 1228

RESULT 12  
O55264  
ID O55264; PRELIMINARY; PRT: 1449 AA.  
AC O55264;

DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE PRECURSOR.  
GN GTF.  
OS Streptococcus salivarius.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1304;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=95122197; PubMed=7822030;  
RX Simpson C.L., Giffard P.M., Jacques N.A.;  
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes  
coding for primer-independent glucosyltransferases.";  
RL Infect. Immun. 63:609-621(1995).  
DR EMBL; L35495; AAC4142.1; -  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 8.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
FT SIGNAL; Transferase.  
FT CHAIN 1 35 POTENTIAL.  
SQ SEQUENCE 1449 AA; 159984 MW; DD62F07306E86A46 CRC64;  
Matches 15; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 TGARTINGOLLYFRANGOVKG 22  
Db 1258 TGLVNINGMLKTFQANGROVKG 1279

RESULT 13  
ID 09L466 PRELIMINARY; PRT; 1477 AA.  
AC 09L466.  
DT 01-OCT-2000 (TREMblrel. 15, Created)  
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
DE DEXTRANSUCRASE (EC 2.4.1.5).  
GN DSR.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=NRRL B-1355;  
RA Arguello-Morales M.A., Remaud-Simeon M., Pizut S., Sarcabal P.,  
RT "Sequence analysis of the gene encoding alternansucrase, a sucrose  
glucosyltransferase from leuconostoc mesenteroides NRRL B-1355.";  
RL Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AJ250112; CAB76565.1; -  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 14.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
DR Transferase; Glycosyltransferase.  
SQ SEQUENCE 1477 AA; 164886 MW; E6F5710DEDFC8831 CRC64;

Query Match 64.3%; Score 72; DB 2; Length 1477;  
Best Local Similarity 68.2%; Pred. No. 0.0075;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGARTINGOLLYFRANGOVKG 22  
Db 1368 TGLVNINGMLKTFQANGROVKG 1389

RESULT 14  
ID 052224 PRELIMINARY; PRT; 1508 AA.  
AC 052224;  
DT 01-JUN-1998 (TREMblrel. 06, Created)  
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE 6-  
DE GLUCOSYLTRANSFERASE).  
GN DSRB.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=NRRL B-1299;  
RA Monchois V., Remaud-Simeon M., Monsan P., Willemot R.M.;  
RL FEMS Microbiol. Lett. 0:0-0(1998).  
CC -1-CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-  
FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).  
DR EMBL; AF030129; AAB95453.1; -  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 14.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
DR Transferase; Glycosyltransferase.  
SQ SEQUENCE 1508 AA; 168511 MW; E70CEB57A70D1F0 CRC64;

Query Match 64.3%; Score 72; DB 2; Length 1508;  
Best Local Similarity 68.2%; Pred. No. 0.0077;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGARTINGOLLYFRANGOVKG 22  
Db 1399 TGLVNINGMLKTFQANGROVKG 1420

RESULT 15  
ID 09EZH5 PRELIMINARY; PRT; 1508 AA.  
AC 09EZH5.  
DT 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
DE DEXTRANSUCRASE; DSRB742.  
GN DSRB742.  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=B-742CB.  
RA Kim H.-S., Kim D., Ryu H.-J., Rodyt J.F.;  
RT "Leuconostoc mesenteroides B-742CB, a dextransucrase gene.";  
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF294469; AAG38021.1; -  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 14.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
SQ SEQUENCE 1508 AA; 168542 MW; E2FCFA0F87AE4F3A CRC64;

Query Match 64.3%; Score 72; DB 2; Length 1508;  
Best Local Similarity 68.2%; Pred. No. 0.0077;  
Matches 15; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGARTINGOLLYFRANGOVKG 22

DB 1399 TGLVNINGNLKRYEQANGROYKG 1420

Search completed: March 27, 2002, 14:26:03  
Job time: 1675 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:49:12 ; Search time 198.55 Seconds

(without alignments)  
7.834 Million cell updates/sec

Title: US-09-290-049a-1

Perfect score: 119

Sequence: 1 ANDHSLTEAMSDNDPPLYLHD 21

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq\_1101.\*  
1: /SID8/gcgdata/geneseq/genesep/AA1980.DAT.\*  
2: /SID8/gcgdata/geneseq/genesep/AA1981.DAT.\*  
3: /SID8/gcgdata/geneseq/genesep/AA1982.DAT.\*  
4: /SID8/gcgdata/geneseq/genesep/AA1983.DAT.\*  
5: /SID8/gcgdata/geneseq/genesep/AA1984.DAT.\*  
6: /SID8/gcgdata/geneseq/genesep/AA1985.DAT.\*  
7: /SID8/gcgdata/geneseq/genesep/AA1986.DAT.\*  
8: /SID8/gcgdata/geneseq/genesep/AA1987.DAT.\*  
9: /SID8/gcgdata/geneseq/genesep/AA1988.DAT.\*  
10: /SID8/gcgdata/geneseq/genesep/AA1989.DAT.\*  
11: /SID8/gcgdata/geneseq/genesep/AA1990.DAT.\*  
12: /SID8/gcgdata/geneseq/genesep/AA1991.DAT.\*  
13: /SID8/gcgdata/geneseq/genesep/AA1992.DAT.\*  
14: /SID8/gcgdata/geneseq/genesep/AA1993.DAT.\*  
15: /SID8/gcgdata/geneseq/genesep/AA1994.DAT.\*  
16: /SID8/gcgdata/geneseq/genesep/AA1995.DAT.\*  
17: /SID8/gcgdata/geneseq/genesep/AA1996.DAT.\*  
18: /SID8/gcgdata/geneseq/genesep/AA1997.DAT.\*  
19: /SID8/gcgdata/geneseq/genesep/AA1998.DAT.\*  
20: /SID8/gcgdata/geneseq/genesep/AA1999.DAT.\*  
21: /SID8/gcgdata/geneseq/genesep/AA2000.DAT.\*  
22: /SID8/gcgdata/geneseq/genesep/AA2001.DAT.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	90.8	1592	14	AA32925
2	63	52.9	1577	17	AA391047
3	60	50.4	2057	21	AA10667
4	51	42.9	486	22	AA36552
5	50	42.0	305	21	AA309845
6	50	42.0	339	21	AA309844
7	48.5	40.8	401	22	AA43544
8	47.5	39.9	534	21	AA167414
9	47	39.5	195	21	AA309941
10	47	39.5	338	15	AA35965
11	46	38.7	118	20	AA560021

12	46	38.7	199	22	AA390817
13	46	38.7	236	20	AA35244
14	45	37.8	336	20	AA23993
15	45	37.8	385	12	AA15238
16	44	37.0	62	21	AA44863
17	44	37.0	161	21	AA328444
18	44	37.0	206	21	AA328443
19	44	37.0	220	22	AA382161
20	44	37.0	384	21	AA329041
21	43	36.1	126	21	AA326416
22	43	36.1	126	21	AA353894
23	43	36.1	170	21	AA326415
24	43	36.1	170	21	AA353893
25	43	36.1	193	21	AA353892
26	43	36.1	194	21	AA326414
27	43	36.1	327	22	AA364817
28	43	36.1	335	20	AA307036
29	43	36.1	335	20	AA305368
30	43	36.1	1195	20	AA375420
31	43	36.1	2466	16	AA371498
32	43	36.1	2466	19	AA375999
33	43	36.1	2466	21	AA390272
34	43	36.1	2485	21	AA319343
35	42.5	35.7	325	22	AA341942
36	42.5	35.7	390	22	AA340156
37	42	35.3	123	22	AA379865
38	42	35.3	174	21	AA360587
39	42	35.3	178	10	AA390061
40	42	35.3	185	21	AA360586
41	42	35.3	189	21	AA360585
42	42	35.3	236	20	AA390021
43	42	35.3	239	20	AA389935
44	42	35.3	259	17	AA383326
45	42	35.3	298	19	AA371486

#### ALIGNMENTS

RESULT 1	
ID	AA32925
AA32925	standard; Protein; 1592 AA.
XX	
AC	AA32925;
XX	
DT	28-JUN-1993 (first entry)
XX	
DE	Glucosyltransferase I.
XX	
KM	GT-1; Streptococcus; dental; carries.
XX	
OS	Streptococcus sobrinus.
XX	
PN	JP05023188-A.
XX	
PD	02-FEB-1993.
XX	
PF	25-JUL-1991; 91JP-0186592.
XX	
PR	25-JUL-1991; 91JP-0186592.
XX	
PA	(FUKU/) FUKU I.
PA	(KATO/) KATO K.
XX	
DR	WPI; 1993-079449/10.
XX	
PT	N-PSDB; AA037760.
PT	DNA sequence glucosyl:transferase-I - comprises Streptococcus
PT	sobrinus DNA sequence with at least one nucleotide added or
XX	deleted
PS	Claim 13; Page 15; 29pp; Japanese.
XX	

C. glutamicum prote  
C. pneumoniae prot  
Amino acid sequenc  
Fusarium oxysporum  
Zea mays protein f  
Arabidopsis thalia  
Arabidopsis thalia  
S. epidermidis ope  
C. militans novel  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Human G3PDH fragme  
Breast cancer asso  
Human HCMV Inducib  
T. thermophilus nit  
Human protein tyro  
Intracellular prot  
Human PRP1 phosph  
Amino acid sequenc  
Human polypeptide  
Human polypeptide  
Corynebacterium gl  
Arabidopsis thalia  
Human Fc gamma rec  
Arabidopsis thalia  
Arabidopsis thalia  
Expressed antigen  
Antigen 1 from glu  
Periplasmic chap  
Helicobacter polyp



OY 1 ANDHLSLEAMSNDPTYLH 20  
||| ||||| 1: 1:;  
Db 665 ankhsiledmgkdpqyv 684

## RESULT 4

AAB36552  
ID AAB36552 standard; Protein: 486 AA.

AC AAB36552;

DT 07-MAR-2001 (first entry)

DE Lawsonia intracellularis flagellar hook protein FlgE SEQ ID NO:1.

KW Lawsonia intracellularis; flgE; flagellar hook protein; vaccine;  
KW intestinal disease; immunogenic; diagnosis; antibacterial; swine;  
KW pig; infection; detection; identification.

OS Lawsonia intracellularis.

MO200069904-A1.

PD 23-NOV-2000.

PE 11-MAY-2000; 2000WO-AU00437.

PR 13-MAY-1999; 99US-0133973.

PA (PFIZ ) PFIZER PROD INC.  
(AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
(PIGR-) PIG RES & DEV CORP.

PI Panaccio M, Rosey EL, Stinista J, Hasse D, Parsons J;  
Ankenbauer RG;

PI MPI; 2001-016210/02.

DR N-PSDB; AAC88037.

XX New immunogenic Lawsonia flgE peptide, its nucleic acid and antibody,  
PT useful in vaccines and diagnosis of Lawsonia infections, particularly  
PT in swine -

PS Claim 13; Page 87-90; 97pp; English.

XX The present sequence is the Lawsonia intracellularis flagellar hook  
CC protein FlgE. The present invention describes an isolated or recombinant  
CC polypeptide (I) that comprises, mimics or cross-reacts with a B- or  
CC T-cell epitope of a FlgE (flagellar hook) polypeptide from a  
CC Lawsonia spp. (I) has antibacterial activity, and induces a specific  
CC humoral immune response. (I) are used as antigens in vaccines to prevent  
CC or treat infection by Lawsonia, in birds and animals, especially pigs,  
CC to raise specific antibodies (Ab) and to detect past or present  
CC infection. Ab are also useful in diagnosis, to detect L. intracellularis  
CC or immunologically cross-reactive species, also for identification of  
CC epitopes in FlgE. Vectors that contain nucleic acids (II) encoding (I)  
CC are also useful in genetic vaccines, and fragments of (II) are useful  
CC as primers or probes for detecting L. intracellularis or related  
CC microorganisms, in hybridisation or amplification assays.

XX Sequence 486 AA;

SO Query Match 42.9%; Score 51; DB 22; Length 486;  
Best Local Similarity 47.1%; Pred. No. 12;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 1 ANDHLSLEAMSNDPT 17  
||| :|||: |||

Db 169 anpyfallawskngt 185

## RESULT 5

AAG09845  
ID AAG09845 standard; Protein: 305 AA.

AC AAG09845;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 7935.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

OS Arabidopsis thaliana.

EP1033405-A2.

PD 06-SEP-2000.

PE 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 18-MAY-1999; 99US-0134370.

PR 19-MAY-1999; 99US-0134768.

PR 20-MAY-1999; 99US-0134941.

PR 21-MAY-1999; 99US-0135124.

PR 24-MAY-1999; 99US-0135353.

PR 25-MAY-1999; 99US-0135629.

PR 27-MAY-1999; 99US-0136021.

PR 28-MAY-1999; 99US-0136392.

PR 01-JUN-1999; 99US-0136782.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137529.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140895.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145213.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147935.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148365.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.

PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 42.0%; Score 50; DB 21; Length 305;  
Best Local Similarity 52.9%; Pred. No. 10;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 3 DHSLEANSNDTPYL 19  
I:|||||I|I|:  
Db 200 dyeglnawskgltlyv 216

RESULT 6  
AAG09844  
ID AAG09844 standard; Protein; 339 AA.  
XX  
AC AAG09844;  
XX  
DT 17-OCT-2000 (first entry)

XX	Arabidopsis thaliana protein fragment SEQ ID NO: 7934.		PR	21-JUN-1999;	99US-0139817.
DE			PR	22-JUN-1999;	99US-0139899;
XX	Protein identification; signal transduction pathway; metabolic pathway;		PR	23-JUN-1999;	99US-0140353.
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		PR	24-JUN-1999;	99US-0140354.
KW	termination sequence.		PR	24-JUN-1999;	99US-0140695.
XX			PR	28-JUN-1999;	99US-0140823.
OS	Arabidopsis thaliana.		PR	30-JUN-1999;	99US-0140991.
XX			PR	30-JUN-1999;	99US-0141287.
PN	EP1033405-A2.		PR	01-JUL-1999;	99US-0141842.
XX			PR	01-JUL-1999;	99US-0142154.
PD	06-SEP-2000.		PR	02-JUL-1999;	99US-0142055.
XX			PR	06-JUL-1999;	99US-0142390.
PF	25-FEB-2000; 2000EP-0301439.		PR	09-JUL-1999;	99US-0142920.
XX			PR	12-JUL-1999;	99US-0142977.
XX			PR	13-JUL-1999;	99US-0143542.
PR	25-FEB-1999;	99US-0121825.	PR	14-JUL-1999;	99US-0143624.
PR	05-MAR-1999;	99US-0123180.	PR	15-JUL-1999;	99US-0144005.
PR	09-MAR-1999;	99US-0123548.	PR	16-JUL-1999;	99US-0144085.
PR	23-MAR-1999;	99US-0125788.	PR	16-JUL-1999;	99US-0144086.
PR	25-MAR-1999;	99US-0126264.	PR	19-JUL-1999;	99US-0144325.
PR	29-MAR-1999;	99US-0126785.	PR	19-JUL-1999;	99US-0144325.
PR	01-APR-1999;	99US-0127462.	PR	19-JUL-1999;	99US-0144331.
PR	06-APR-1999;	99US-0128234.	PR	19-JUL-1999;	99US-0144332.
PR	08-APR-1999;	99US-0128714.	PR	19-JUL-1999;	99US-0144333.
PR	16-APR-1999;	99US-0129845.	PR	19-JUL-1999;	99US-0144334.
PR	19-APR-1999;	99US-0130047.	PR	19-JUL-1999;	99US-0144335.
PR	21-APR-1999;	99US-0130449.	PR	20-JUL-1999;	99US-0144352.
PR	23-APR-1999;	99US-0130510.	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	99US-0130891.	PR	20-JUL-1999;	99US-0144684.
PR	28-APR-1999;	99US-0131449.	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0145088.
PR	30-APR-1999;	99US-0132407.	PR	21-JUL-1999;	99US-0145088.
PR	04-MAY-1999;	99US-0132484.	PR	22-JUL-1999;	99US-0145085.
PR	05-MAY-1999;	99US-0132485.	PR	22-JUL-1999;	99US-0145087.
PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145089.
PR	07-MAY-1999;	99US-0132487.	PR	22-JUL-1999;	99US-0145192.
PR	11-MAY-1999;	99US-0132863.	PR	23-JUL-1999;	99US-0145145.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145524.
PR	14-MAY-1999;	99US-0134219.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134221.	PR	27-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145918.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	99US-0134941.	PR	28-JUL-1999;	99US-0145951.
PR	20-MAY-1999;	99US-0135124.	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	99US-0135623.	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136392.	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139452.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139455.	PR	16-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149178.
PR	18-JUN-1999;	99US-0139456.	PR	17-AUG-1999;	99US-0149178.
PR	18-JUN-1999;	99US-0139457.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139462.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139463.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139750.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.

PR 27-AUG-1999: 99US-0151080.  
PR 30-AUG-1999: 99US-0151303.  
PR 31-AUG-1999: 99US-0151338.  
PR 01-SEP-1999: 99US-0151930.  
PR 07-SEP-1999: 99US-0152363.  
PR 10-SEP-1999: 99US-0153070.  
PR 13-SEP-1999: 99US-0153758.  
PR 15-SEP-1999: 99US-0154018.  
PR 16-SEP-1999: 99US-0154039.  
PR 20-SEP-1999: 99US-0154779.  
PR 22-SEP-1999: 99US-0155139.  
PR 23-SEP-1999: 99US-0155486.  
PR 24-SEP-1999: 99US-0155659.  
PR 28-SEP-1999: 99US-0156458.  
PR 29-SEP-1999: 99US-0156596.  
PR 04-OCT-1999: 99US-0157117.  
PR 05-OCT-1999: 99US-0157753.  
PR 06-OCT-1999: 99US-0157865.  
PR 07-OCT-1999: 99US-0158029.  
PR 08-OCT-1999: 99US-0158232.  
PR 12-OCT-1999: 99US-0158369.  
PR 13-OCT-1999: 99US-0159293.  
PR 13-OCT-1999: 99US-0159294.  
PR 13-OCT-1999: 99US-0159295.  
PR 14-OCT-1999: 99US-0159329.  
PR 14-OCT-1999: 99US-0159330.  
PR 14-OCT-1999: 99US-0159331.  
PR 14-OCT-1999: 99US-0159637.  
PR 14-OCT-1999: 99US-0159637.  
PR 18-OCT-1999: 99US-0159584.  
PR 21-OCT-1999: 99US-0160741.  
PR 21-OCT-1999: 99US-0160767.  
PR 21-OCT-1999: 99US-0160768.  
PR 21-OCT-1999: 99US-0160770.  
PR 21-OCT-1999: 99US-0160814.  
PR 21-OCT-1999: 99US-0160815.  
PR 22-OCT-1999: 99US-0160980.  
PR 22-OCT-1999: 99US-0160981.  
PR 22-OCT-1999: 99US-0160989.  
PR 25-OCT-1999: 99US-0161404.  
PR 25-OCT-1999: 99US-0161405.  
PR 25-OCT-1999: 99US-0161406.  
PR 26-OCT-1999: 99US-0161359.  
PR 26-OCT-1999: 99US-0161360.  
PR 26-OCT-1999: 99US-0161361.  
PR 28-OCT-1999: 99US-0161820.  
PR 28-OCT-1999: 99US-0161892.  
PR 28-OCT-1999: 99US-0161993.  
PR 29-OCT-1999: 99US-0162142.  
PR 29-OCT-1999: 99US-0162142.

Query Match 42.0%; Score 50; DB 21; Length 339;  
Best Local Similarity 52.9%; Pred. No. 12;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 DHSLEAFMSDNDPYL 19  
|:|:|:|:|:|:|:|:|:|:|:  
Db 234 dyegimeawskdgtly 250

RESULT 7  
ID AAM43544  
AC AAM43544: standard; Protein; 401 AA.  
XX  
DE 22-OCT-2001 (first entry)  
XX  
XX Human polypeptide SEQ ID NO 222.  
KW Human; antitarrhritic; antirheumatic; antiproliferative; vasotropic;  
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;  
KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;  
KW neuroprotective; antiallergic; hepatotropic; antidiabetic;

KW antiinflammatory; antitumor; anticonvulsant; antibacterial;  
KW antiparasitic; cardiant; gene therapy; cancer; immune disorder;  
KW cardiovascular disorder; neurological disease; infection; human.  
OS Homo sapiens.  
XX  
XX WO200155308-A2.  
PN  
XX  
PD 02-AUG-2001.  
XX  
PE 17-JAN-2001; 2001WO-US01309.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225457.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0232981.  
PR 14-SEP-2000; 2000US-0232987.  
PR 14-SEP-2000; 2000US-0232997.  
PR 14-SEP-2000; 2000US-0232999.  
PR 14-SEP-2000; 2000US-0232999.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0251989.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.

PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-488781/53.  
 DR N-PSDB; AAI63850.  
 XX  
 PT New isolated nucleic acids and polypeptides, useful for diagnosing,  
 PT treating and/or preventing human diseases and disorders -  
 XX  
 PS Claim 11; SEQ ID NO 222; 664pp + Sequence Listing: English.  
 XX  
 CC The invention relates to human polynucleotides (AAI63803-AAI64012) and  
 CC the encoded proteins (AAI634497-AAI63660) useful for preventing, treating  
 CC or ameliorating medical conditions e.g. by protein or gene therapy. The  
 CC genes were isolated from a range of human tissues disclosed in the  
 CC specification. The nucleic acids, proteins, antibodies and (ant)agonists  
 CC are useful in the diagnosis, treatment and prevention of: (a) cancer,  
 CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,  
 CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or  
 CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,  
 CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 401 AA;  
 Query Match 40.8%; Score 48.5; DB 22; Length 401;  
 Best Local Similarity 47.4%; Pred. No. 24;  
 Matches 9; Conservative 3; Mismatches 4; Indels 3; Gaps 1;  
 QY 5 LSIEMWSDN--DTPYLH 20  
 Db 46 valleywkdntdpxh 64  
 ::||| || | |||  
 RESULT 8  
 ID AAI67414 standard; protein; 534 AA.  
 XX  
 AC AAI67414;  
 XX  
 DT 12-MAY-2000 (first entry)  
 XX  
 DE Arabidopsis aldehyde dehydrogenase (ALDH)-3.  
 XX  
 KW Arabidopsis; plant plastid; acetyl CoA synthetase; E3 subunit; ACS;  
 KW plant plastidic pyruvate dehydrogenase; pPDH; ATP citrate lyase; ACL;  
 KW pyruvate decarboxylase; pDC; aldehyde dehydrogenase; ALDH; acetyl CoA;  
 KW fatty acid; flavonoid; enzyme; phytochemical; pyruvate decarboxylase;  
 KW acetyl CoA hydrolase; mitochondrial pyruvate dehydrogenase.  
 XX  
 OS Arabidopsis sp.  
 XX  
 PN W020000619-A2.  
 PN  
 PD 06-JAN-2000.  
 PD  
 XX 25-JUN-1999; 99WO-US14382.  
 PF  
 XX 26-JUN-1998; 98US-0090717.  
 PR  
 XX (IOWA ) UNIV IOWA STATE RES FOUND INC.

Query Match	Best Local Similarity	Score	DB	Length
Matches	9: Conservative	52.9%	2: Mismatches	5: Indels
				1: Gaps
QY	2 NDHLSILEAWSNDNTPY 18			
DB	132 ndeiaaleltw-dngkpy 147			
RESULT	9			
AAAG09941				
ID	AAAG09941 standard; Protein: 195 AA.			
AAAG09941:				
XX	17-OCT-2000 (first entry)			
DT	Arabidopsis thaliana protein fragment SEQ ID NO: 8067.			
XX				
DE	Protein identification; signal transduction pathway; metabolic pathway;			
KW	hybridisation assay; genetic mapping; gene expression control; promoter;			
RN	termination sequence.			
XX				
OS	Arabidopsis thaliana.			
XX				
PN	EP1033405-A2.			
PD	06-SEP-2000.			
XX				
PF	25-FEB-2000; 2000EP-0301439.			
XX				
PR	25-FEB-1999; 990US-0121825.			
PR	05-MAR-1999; 990US-0123180.			
PR	09-MAR-1999; 990US-0123548.			
PR	23-MAR-1999; 990US-0125788.			
PR	25-MAR-1999; 990US-0126264.			
PR	29-MAR-1999; 990US-0126785.			
PR	01-APR-1999; 990US-0127462.			
PR	06-APR-1999; 990US-0128234.			

PR	18-APR-1999;	9905-0128771;
PR	16-APR-1999;	9905-0128685;
PR	19-APR-1999;	9905-0130077;
PR	21-APR-1999;	9905-0130449;
PR	23-APR-1999;	9905-0130510;
PR	28-APR-1999;	9905-0130891;
PR	30-APR-1999;	9905-0132048;
PR	30-APR-1999;	9905-0133407;
PR	04-MAY-1999;	9905-0133484;
PR	05-MAY-1999;	9905-0133485;
PR	06-MAY-1999;	9905-0133486;
PR	06-MAY-1999;	9905-0133487;
PR	07-MAY-1999;	9905-0132863;
PR	11-MAY-1999;	9905-0134256;
PR	14-MAY-1999;	9905-0134218;
PR	14-MAY-1999;	9905-0134219;
PR	14-MAY-1999;	9905-0134321;
PR	14-MAY-1999;	9905-0134370;
PR	18-MAY-1999;	9905-0134941;
PR	19-MAY-1999;	9905-0135124;
PR	20-MAY-1999;	9905-0135353;
PR	21-MAY-1999;	9905-0135359;
PR	24-MAY-1999;	9905-0135629;
PR	25-MAY-1999;	9905-0136021;
PR	27-MAY-1999;	9905-0136392;
PR	28-MAY-1999;	9905-0136782;
PR	01-JUN-1999;	9905-0137222;
PR	03-JUN-1999;	9905-0137528;
PR	04-JUN-1999;	9905-0137502;
PR	07-JUN-1999;	9905-0137724;
PR	08-JUN-1999;	9905-0138094;
PR	10-JUN-1999;	9905-0138350;
PR	10-JUN-1999;	9905-0138847;
PR	14-JUN-1999;	9905-0139311;
PR	16-JUN-1999;	9905-0139452;
PR	16-JUN-1999;	9905-0139453;
PR	17-JUN-1999;	9905-0139492;
PR	18-JUN-1999;	9905-0139454;
PR	18-JUN-1999;	9905-0139455;
PR	18-JUN-1999;	9905-0139456;
PR	18-JUN-1999;	9905-0139457;
PR	18-JUN-1999;	9905-0139458;
PR	18-JUN-1999;	9905-0139459;
PR	18-JUN-1999;	9905-0139460;
PR	18-JUN-1999;	9905-0139461;
PR	18-JUN-1999;	9905-0139462;
PR	18-JUN-1999;	9905-0139463;
PR	18-JUN-1999;	9905-0139464;
PR	18-JUN-1999;	9905-0139465;
PR	18-JUN-1999;	9905-0139466;
PR	18-JUN-1999;	9905-0139467;
PR	18-JUN-1999;	9905-0139468;
PR	18-JUN-1999;	9905-0139469;
PR	18-JUN-1999;	9905-0139470;
PR	18-JUN-1999;	9905-0139471;
PR	18-JUN-1999;	9905-0139472;
PR	18-JUN-1999;	9905-0139473;
PR	18-JUN-1999;	9905-0139474;
PR	18-JUN-1999;	9905-0139475;
PR	18-JUN-1999;	9905-0139476;
PR	18-JUN-1999;	9905-0139477;
PR	18-JUN-1999;	9905-0139478;
PR	18-JUN-1999;	9905-0139479;
PR	18-JUN-1999;	9905-0139480;
PR	18-JUN-1999;	9905-0139481;
PR	18-JUN-1999;	9905-0139482;
PR	18-JUN-1999;	9905-0139483;
PR	18-JUN-1999;	9905-0139484;
PR	18-JUN-1999;	9905-0139485;
PR	18-JUN-1999;	9905-0139486;
PR	18-JUN-1999;	9905-0139487;
PR	18-JUN-1999;	9905-0139488;
PR	18-JUN-1999;	9905-0139489;
PR	18-JUN-1999;	9905-0139490;
PR	18-JUN-1999;	9905-0139491;
PR	18-JUN-1999;	9905-0139492;
PR	18-JUN-1999;	9905-0139493;
PR	18-JUN-1999;	9905-0139494;
PR	18-JUN-1999;	9905-0139495;
PR	18-JUN-1999;	9905-0139496;
PR	18-JUN-1999;	9905-0139497;
PR	18-JUN-1999;	9905-0139498;
PR	18-JUN-1999;	9905-0139499;
PR	18-JUN-1999;	9905-0139500;
PR	18-JUN-1999;	9905-0139501;
PR	18-JUN-1999;	9905-0139502;
PR	18-JUN-1999;	9905-0139503;
PR	18-JUN-1999;	9905-0139504;
PR	18-JUN-1999;	9905-0139505;
PR	18-JUN-1999;	9905-0139506;
PR	18-JUN-1999;	9905-0139507;
PR	18-JUN-1999;	9905-0139508;
PR	18-JUN-1999;	9905-0139509;
PR	18-JUN-1999;	9905-0139510;
PR	18-JUN-1999;	9905-0139511;
PR	18-JUN-1999;	9905-0139512;
PR	18-JUN-1999;	9905-0139513;
PR	18-JUN-1999;	9905-0139514;
PR	18-JUN-1999;	9905-0139515;
PR	18-JUN-1999;	990

[illegible]





PT also new glyceraldehyde-3-phosphate dehydrogenase gene.  
XX  
PS Disclosure; Fig.15; 82pp; German.  
XX  
CC A cosmid bank of T. niveum ATCC 34921 genomic DNA was screened with  
CC a fragment of the glyceraldehyde-3-phosphate-dehydrogenase (GAPDH)  
CC gene from *Pericillium chrysogenum*. A 2.2 kb fragment was  
CC identified (AA078277) and cloned in pUC18 to give pGTL. A 665 bp  
CC HindIII-HindII fragment of this was used to screen cosmid DNA to  
CC identify a DNA fragment (AA078279) containing the GAPDH promoter.  
CC A sequence (AA078278) encoding the GAPDH gene (gpdA) was isolated from  
CC a cDNA library. The full sequence of the genomic gpdA structural  
CC gene is given in AA078280 and the corresponding protein sequence in  
CC AA065965. The GAPDH promoter is useful for expressing foreign genes  
CC in T. niveum.  
XX  
SO Sequence 338 AA:  
Query Match 39.5%; Score 47; DB 15; Length 338;  
Best Local Similarity 47.1%; Pred. No. 33;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
QY 2 NDHLSLEAWSNDTPY 18  
Db 304 ndhfaklvswydwngy 320  
||| : | : | : |  
| : | : | : |  
RESULT 11  
AA060021  
ID AA060021 standard; Protein: 118 AA.  
XX  
AC AA060021;  
XX  
DT 31-JAN-2000 (first entry)  
XX  
DE Human endometrium tumour EST encoded protein 81.  
XX  
DE Endometrium; human; tumour; cancer; anticancer; cytostatic; EST:  
KW treatment; uterine; gene therapy; expressed sequence tag.  
XX  
OS Homo sapiens.  
XX  
PN DEJ9817948-A1.  
XX  
PD 21-OCT-1999.  
XX  
PR 17-APR-1998; 98DE-1017948.  
XX  
PR 17-APR-1998; 98DE-1017948.  
XX  
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.  
XX  
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
XX  
XX WPI: 1999-591957/51.  
DR N-PSDB; AA242007.  
XX  
PT New nucleic acid sequences expressed in uterine cancer tissues, and  
PT derived polypeptides, for treatment of uterine and endometrial cancer  
PT and identification of therapeutic agents  
XX  
PS Claim 23; Page 307; 444pp; German.  
XX  
CC This invention describes novel human nucleic acid (cDNA) sequences (A),  
CC that are highly expressed in uterine tumour tissue and which have  
CC anticancer and cytostatic activity. (A) are used (i) for recombinant  
CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)  
CC are used (i) to identify agents suitable for treatment of uterine or  
CC endometrial cancer, (ii) directly for treating these forms of cancer  
CC (including expression from gene therapy vectors) and (iii) for  
CC generation of specific antibodies. (A) are identified by assembling ESTs  
CC (expressed sequence tags) from a particular tissue type before comparison

CC of expression patterns. This allows a significantly longer fragment of  
CC the gene to be revealed, so should reduce the number of failures  
CC associated with the fact that ESTs from different libraries may represent  
CC different parts of the same unknown gene, distorting the estimated  
CC frequency of occurrence in a particular tissue. AA059941-Y60328 represent  
CC protein fragments encoded by the human endometrium tumour cDNA library  
CC derived EST fragments represented in AA241981-242121.  
XX  
SO Sequence 118 AA:  
Query Match 38.7%; Score 46; DB 20; Length 118;  
Best Local Similarity 50.0%; Pred. No. 14;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;  
QY 3 DNLSTLEAWSNDTPY 20  
Db 23 dpldlldagqdkdtpssh 40  
| : | : | : |  
| : | : | : |  
RESULT 12  
AA090817  
ID AA090817 standard; Protein: 199 AA.  
XX  
AC AA090817;  
XX  
DT 26-SEP-2001 (first entry)  
XX  
DE C glutamicum protein fragment SEQ ID NO: 4571.  
XX  
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN EP1108790-A2.  
XX  
PD 20-JUN-2001.  
XX  
PF 18-DEC-2000; 2000EP-0127688.  
XX  
PR 16-DEC-1999; 99JP-0377484.  
PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.  
XX  
PA (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
XX  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX  
XX WPI: 2001-376931/40.  
DR N-PSDB; AA066036.  
XX  
PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
PT  
XX  
PS Claim 17; SEQ ID NO: 4571; 246pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium *Corynebacterium glutamicum*. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived  
CC from coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a protein described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.

XX Sequence 199 AA;

Query Match 38.7%; Score 46; DB 22; Length 199;  
Best Local Similarity 41.2%; Pred. No. 25;  
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 2 NDHLSILEAMSDNDPPY 18  
:|::|:|:|:|  
Db 172 ddsqdlvewlqdlrly 188

RESULT 13  
AAV35244  
ID AAV35244 standard; Protein: 336 AA.

XX AAV35244;

XX 13-SEP-1999 (first entry)

XX C. pneumoniae protein involved in intermediate metabolism.

KM Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

PN WO9927105-A2.

PD 03-JUN-1999.

XX 20-NOV-1998; 98WO-1B01890.

XX 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

XX (GEST ) GENSET.

PI Griffiths R;

DR WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae

AAV34584-Y35879 represent the proteins encoded by all the open reading frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.

C. pneumoniae causes respiratory disease such as pneumonia and bronchitis and is thought to be a contributing factor in heart disease, sarcoidosis, sinusitis, purulent otitis media, erythema nodosum or pharyngitis. The polypeptides encoded by the open reading frames of the C. pneumoniae genome (see AAV34584-Y35879) can be used in immunogenic compositions as vaccines. Vectors containing C. pneumoniae nucleotide sequences can also be used as immunogenic compositions, especially where the vector directs the expression of a neutralising epitope of C. pneumoniae.

XX Sequence 336 AA;

Query Match 38.7%; Score 46; DB 20; Length 336;  
Best Local Similarity 52.9%; Pred. No. 46;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 2 NDHLSILEAMSDNDPPY 18  
|||:|:|:|:|  
Db 304 ndrftklvawydnetcy 320

RESULT 14

AAV23993  
ID AAV23993 standard; Protein: 236 AA.

XX AAV23993;

DT 22-SEP-1999 (first entry)

DE Amino acid sequence of a human p16 protein.

KW p53 transcription tag; p53 status; cancer; cytotoxicity; carcinogenicity;  
KW neoplastic; p16.

XX Homo sapiens.

PN WO9914356-A2.

PD 25-MAR-1999.

XX 17-SEP-1998; 98WO-US19300.

PR 30-MAR-1998; 98US-0079817.

PR 17-SEP-1997; 97US-0059153.

XX (UYJO ) UNIV JOHNS HOPKINS.

PI Kinzler KW, Polyak K, Vogelstein B;

DR WPI; 1999-443793/37.

XX Use of p53 transcription tags to determine p53 status in, e.g.  
cancer diagnosis

PS Claim 39; Page 67; 73pp; English.

The specification describes the use of p53 transcription tags for developing products to determine p53 status, to diagnose cancer and to evaluate cytotoxicity or carcinogenicity of a test agent.  
CC A method for diagnosing cancer or determining p53 status in a sample suspected for being neoplastic comprises comparing the level of transcription of an RNA transcript in a first sample (s1) of a first tissue (t1) to the level of transcription of the transcript in a second sample (s2) of a second tissue (s2), where s1 is suspected of being neoplastic and s2 is a normal human tissue (of the same type) and the transcript is identified by a tag; and categorizing s1 as neoplastic or as having a mutant p53 when transcription is found to be the same or lower in the first, than in s2. The methods and products can be used to determine p53 status, to diagnose cancer and to evaluate cytotoxicity or carcinogenicity of a test agent. AAV23992-Y24000 represent human p16 proteins.

XX Sequence 236 AA;

Query Match 37.8%; Score 45; DB 20; Length 236;  
Best Local Similarity 44.0%; Pred. No. 44;  
Matches 11; Conservative 3; Mismatches 3; Indels 8; Gaps 2;

OY 1 ANDHLSIL-----EAMSDNDPPYLH 20  
||:|:|:|:|:|:|:|:|  
Db 30 adthrlilgkpwmerw---etpymh 51

RESULT 15

AAV15238  
ID AAV15238 standard; Protein: 385 AA.

XX AAV15238;

DT 11-FEB-1992 (first entry)

XX Fusarium oxysporum F-family cellulohydrolase.

KW cellulose; carbohydrate binding domain; CBD; fungi; cellulase.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:59:38 ; Search time 53.4 Seconds

(without alignments)  
14.419 Million cell updates/sec

Title: US-09-290-049a-1  
Perfect score: 119  
Sequence: 1 ANDHLILBAMSNDPRYLHD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	119	100.0	1476 1	GTFB_STRMU
2	110	92.4	1375 1	GTFB_STRMU
3	108	90.8	1592 1	GTFB_STRMU
4	108	90.8	1597 1	GTFB_STRMU
5	78	65.5	1365 1	GTFB_STRMU
6	75	63.0	1462 1	GTFB_STRMU
7	57	47.9	330 1	GTFB_STRMU
8	49	41.2	336 1	GTFB_STRMU
9	47.5	39.9	330 1	GTFB_STRMU
10	46	38.7	275 1	GTFB_STRMU
11	46	38.7	335 1	GTFB_STRMU
12	46	38.7	476 1	GTFB_STRMU
13	45	37.8	236 1	GTFB_STRMU
14	45	37.8	385 1	GTFB_STRMU
15	44.5	37.4	470 1	GTFB_STRMU
16	44	37.0	234 1	GTFB_STRMU
17	44	37.0	330 1	GTFB_STRMU
18	44	37.0	332 1	GTFB_STRMU
19	44	37.0	332 1	GTFB_STRMU
20	44	37.0	337 1	GTFB_STRMU
21	44	37.0	337 1	GTFB_STRMU
22	44	37.0	337 1	GTFB_STRMU
23	44	37.0	337 1	GTFB_STRMU
24	44	37.0	337 1	GTFB_STRMU
25	44	37.0	337 1	GTFB_STRMU
26	43.5	36.6	115 1	GTFB_STRMU
27	43.5	36.6	130 1	GTFB_STRMU
28	43.5	36.6	130 1	GTFB_STRMU
29	43.5	36.6	130 1	GTFB_STRMU
30	43.5	36.6	130 1	GTFB_STRMU
31	43	36.1	235 1	GTFB_STRMU
32	43	36.1	236 1	GTFB_STRMU
33	43	36.1	320 1	GTFB_STRMU

34	43	36.1	322 1	G3P_SHEEP
35	43	36.1	332 1	G3P_PIG
36	43	36.1	332 1	G3P_RABIT
37	43	36.1	334 1	G3P2_HUMAN
38	43	36.1	337 1	G3P_MYGE
39	43	36.1	337 1	G3P_MYGE
40	43	36.1	424 1	IVD_RAT
41	43	36.1	674 1	COA_BACTJ
42	42.5	35.7	2485 1	PTND_HUMAN
43	42.5	35.7	441 1	HGD_STRCO
44	42.5	35.7	491 1	GPD_ECOLI
45	42.5	35.7	645 1	IME2_YEAST
			764 1	Y157_YEAST

## ALIGNMENTS

RESULT 1  
ID GTFB\_STRMU STANDARD: PRT: 1476 AA.  
AC P08987: 069381; 069384; 069387; 069390; 069396;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)  
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
DE GTFB.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-GS-5;  
RX MEDLINE=87308013; PubMed=3040685;  
RA Shiroza T., Ueda S., Kuramitsu H.K.;  
RT \*Sequence analysis of the gtf gene from Streptococcus mutans.\*;  
RL J. Bacteriol. 169:4263-4270(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT4239, MT4245, MT4251, MT4467, AND MT8148;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,  
RA Kimura S., Hamada S.;  
RT \*Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans.\*;  
RL FEMS Microbiol. Lett. 161:331-336(1998).  
CC - FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC - CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).  
CC - D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).  
CC - SURCELLULAR LOCATION: SECRETED.  
CC - DISEASE: DENTAL CARIES.  
CC - MISCELLANEOUS: GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF GLUCANS.  
CC - SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-BINDING PROTEIN FROM S.MUTANS.  
CC - BINDING PROTEIN FROM S.MUTANS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: M17361; AAA8588.1; -  
CC EMBL: D88651; BAA26101.1; -  
CC EMBL: D88654; BAA26105.1; -  
DR

```

DR EMBL: D86557; BAA26109.1; -
DR EMBL: D86600; BAA26113.1; -
DR EMBL: D89977; BAA26119.1; -
DR PIR: B3135; B3135.
DR InterPro: IPR002479; CW_binding_70.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 13.
DR Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.
KW SIGNAL.
FT CHAIN 1 34
FT DOMAIN 35 1476
FT DOMAIN 1097 1476
FT REPEAT 1097 1130
FT REPEAT 1161 1470
FT REPEAT 1161 1210
FT REPEAT 1225 1275
FT REPEAT 1290 1340
FT REPEAT 1355 1405
FT REPEAT 1420 1470
FT REPEAT 62 62
FT VARIANT 62 65
FT VARIANT 65 65
FT VARIANT 65 65
FT VARIANT 68 68
FT VARIANT 78 78
FT VARIANT 86 86
FT VARIANT 89 89
FT VARIANT 168 168
FT VARIANT 276 276
FT VARIANT 399 399
FT VARIANT 474 474
FT VARIANT 512 512
FT VARIANT 519 519
FT VARIANT 701 701
FT VARIANT 708 708
FT VARIANT 938 938
FT VARIANT 952 957
FT VARIANT 963 964
FT VARIANT 968 970
FT VARIANT 1086 1086
FT VARIANT 1158 1158
FT VARIANT 1163 1163
FT VARIANT 1168 1168
FT VARIANT 1182 1182
FT VARIANT 1234 1234
FT VARIANT 1263 1263
FT VARIANT 1263 1263
FT VARIANT 1364 1364
FT VARIANT 1372 1372
FT VARIANT 1329 1329
FT VARIANT 1394 1394
FT VARIANT 1402 1402
FT VARIANT 1459 1459
FT CONFLICT 570 570
FT CONFLICT 800 817
FT CONFLICT 1310 1310
FT SEQUENCE 1476 AA; 165685 MW; 3479862807694D98 CRC64;

Query Match 100.0%; Score 119; DB 1; Length 1476;
Best Local Similarity 100.0%; Pred. No. 2,1e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ANDHSLTEAMSDNDTPYLHD 21
481 ANDHSLTEAMSDNDTPYLHD 501

```

```

RESULT 2
ID GTFIC_STRMU STANDARD; PRT; 1375 AA.
AC GTFIC_STRMU
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)
DE (DEKTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTFIC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBL_Taxid=1309;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-GS-5;
RX MEDLINE-89137980; PubMed-2976010;
RX Ueda S., Shiroza T., Kuramitsu H.K.;
RT "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.";
RL gene 69:101-109(1988).
RP SEQUENCE OF 1-349 FROM N.A.
RX STRAIN-GS-5;
RX MEDLINE-87308013; PubMed-3040685;
RX Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans GS-5.";
RL J. Bacteriol. 169:4263-4270(1987).
RP -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDATE THE
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRITS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) -
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
1,3-LINKED GLUCOSE AND SOME 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
FORMS OF GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
BINDING PROTEIN FROM S. MUTANS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M22054; AAA88592.1; -
DR EMBL: M17361; AAA88589.1; -
DR PIR: JTO345; JTO345.
DR PIR: C3135; C3135.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 7.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 34
FT CHAIN 35 1375
FT DOMAIN 35 1050
FT DOMAIN 1126 1375
FT DOMAIN 1126 1375
FT REPEAT 1126 1159
FT REPEAT 1169 1200
FT REPEAT 1227 1238
FT REPEAT 1253 1303
FT REPEAT 1318 1330
FT REPEAT 1375 AA; 153022 MW; D48B0CBEOACEL3 CRC64;

```

Query Match 92.4%; Score 110; DB 1; Length 1375;  
 Best Local Similarity 95.2%; Pred. No. 4.7e-09;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ANDHLSLEAMSDNDPTPLHD 21  
 |||||  
 DB 507 ANDHLSLEAMSYNDPTPLHD 527

RESULT 3  
 GTF2\_STRDO STANDARD; PRT; 1592 AA.  
 ID GTF2\_STRDO

DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 Streptococcus downei (Streptococcus sobrinus).  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 Streptococcus.  
 NCBI\_TaxID=1317;  
 ON NCBI [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=6715;  
 RX MEDLINE=9112327; PubMed=1704006;  
 RA Abo H., Matsumura T., Kodama T., Ohba H., Fukui K., Kato K.,  
 Kagawa H.;  
 RT "Peptide sequences for sucrose splitting and glucan binding within  
 streptococcus sobrinus glucosyltransferase (water-insoluble glucan  
 synthetase).";  
 RL J. Bacteriol. 173:989-996(1991).  
 CC -I- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDATE THE  
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -I- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -I- SUBCELLULAR LOCATION: SECRETED.  
 CC -I- DISEASE: DENTAL CARIES.  
 CC -I- MISCELLANEOUS: GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
 WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
 FORMS OF GLUCANS.  
 CC -I- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 BINDING PROTEIN FROM S.MUTANS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: D90213; BAA14241.1; -  
 CC PTR: A38175; A38175.  
 DR HSP; P00695; ZHE.  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 16.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.  
 FT SIGNAL 1 38 POTENTIAL.  
 FT CHAIN 39 1592 GLUCOSYLTRANSFERASE-1.  
 FT DOMAIN 39 1044 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1093 1592 GLUCAN-BINDING (APPROXIMATE).  
 FT DOMAIN 1093 1592 6.5 X TANDEM REPEATS.  
 FT REPEAT 1093 1142 1.  
 FT REPEAT 1158 1207 2.  
 FT REPEAT 1222 1272 3.

FT REPEAT 1287 1337 4.  
 FT REPEAT 1402 1451 5.  
 FT REPEAT 1514 1563 6.  
 FT REPEAT 1577 1592 7 (INCOMPLETE).  
 SQ SEQUENCE 1592 AA; 176167 MW; BCOA66D07935IECF CRC64;

Query Match 90.8%; Score 108; DB 1; Length 1592;  
 Best Local Similarity 85.7%; Pred. No. 1.1e-08;  
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSLEAMSDNDPTPLHD 21  
 ||:|||||  
 DB 477 ANNHVSIVEAMSDNDPTPLHD 497

RESULT 4  
 GTF1\_STRDO STANDARD; PRT; 1597 AA.  
 ID GTF1\_STRDO

DT 01-JUL-1988 (Rel. 11, Created)  
 DT 01-JUL-1988 (Rel. 11, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 Streptococcus downei (Streptococcus sobrinus).  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 Streptococcus.  
 NCBI\_TaxID=1317;  
 ON NCBI [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=MFE28;  
 RX MEDLINE=87308014; PubMed=3040686;  
 RA Ferretti V.J., Gilpin M.L., Russell R.R.B.;  
 RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus  
 sobrinus MFE28.";  
 RL J. Bacteriol. 169:4271-4278(1987).  
 CC -I- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDATE THE  
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -I- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -I- SUBCELLULAR LOCATION: SECRETED.  
 CC -I- DISEASE: DENTAL CARIES.  
 CC -I- MISCELLANEOUS: GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
 WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
 FORMS OF GLUCANS.  
 CC -I- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 BINDING PROTEIN FROM S.MUTANS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: M17391; AAC63063.1; -  
 CC InterPro: IPR002479; CW\_binding.  
 CC InterPro: IPR003318; Glyco\_hydro\_70.  
 CC Pfam: PF01473; CW\_binding\_1; 19.  
 CC Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.  
 FT SIGNAL 1 38 POTENTIAL.  
 FT CHAIN 39 1597 GLUCOSYLTRANSFERASE-1.  
 FT DOMAIN 39 1050 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1099 1597 GLUCAN-BINDING (APPROXIMATE).  
 FT DOMAIN 1099 1597 1.25 A, 2 B, AND 5 AC REPEATS.  
 FT REPEAT 1099 1132 A REPEAT.

FT REPEAT 1163 1213 AC REPEAT.  
FT REPEAT 1227 1277 AC REPEAT.  
FT REPEAT 1292 1342 AC REPEAT.  
FT REPEAT 1352 1399 B REPEAT.  
FT REPEAT 1406 1455 AC REPEAT.  
FT REPEAT 1465 1512 B REPEAT.  
FT REPEAT 1519 1568 AC REPEAT.  
FT REPEAT 1582 1597 A REPEAT (INCOMPLETE).  
SQ SEQUENCE 1597 AA; 177080 MW; B9E6A20068798E CRC64;

Query Match 90.8%; Score 108; DB 1; Length 1597;  
Best Local Similarity 85.7%; Pred. No. 1.2e-08;  
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHLSILEMSDNDPPLYLD 21  
DB 483 ANNHVSIVEAWSNDPPLYLD 503

RESULT 5  
GTFD\_STRDO STANDARD; PRT; 1365 AA.  
ID GTFD\_STRDO  
AC P29336;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
GN GTF-S.  
OS Streptococcus downei (Streptococcus sobrinus).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1317;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MFE28;  
RX MEDLINE=90316665; PubMed=2142479;  
RA Gilmore K.S., Russell R.R., Ferretti J.J.;  
RT "Analysis of the Streptococcus downei gtfS gene, which specifies a  
glucosyltransferase that synthesizes soluble glucans.";  
RL Infect. Immun. 58:2452-2458(1990).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
CC -1- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF  
PRIMER GLUCAN UNLIKE GTF-I.  
CC -1- DISEASE: DENTAL CARIES.  
CC -1- MISCELLANEOUS: GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA  
1,6-GLUCOSE).  
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
BINDING PROTEIN FROM S. MUTANS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M30943; AAA26898.1; .  
CC PTR: A41483; A41483.  
CC InterPro: IPR002479; CW\_binding.  
CC InterPro: IPR003318; Glyco\_hydro\_70.  
CC Pfam: PF02324; CW\_binding\_1; 10.  
CC Transferrase; Glyco\_hydro\_70; 1.  
KW TRANSFERRASE; GLUCOSYLTRANSFERASE; Signal; Repeat; Dental caries.  
FT SIGNAL 1 36  
FT CHAIN 37 1365 GLUCOSYLTRANSFERASE-S.

FT DOMAIN 37 1050 CATALYTIC (APPROXIMATE).  
FT DOMAIN 1083 1365 GLUCAN-BINDING (APPROXIMATE).  
FT DOMAIN 1083 1365 4.5 X TANDEM REPEATS.  
FT REPEAT 1083 1131 1.  
FT REPEAT 1150 1199 2.  
FT REPEAT 1225 1274 3.  
FT REPEAT 1289 1339 4.  
FT REPEAT 1353 1365 5 (INCOMPLETE).  
SQ SEQUENCE 1365 AA; 151590 MW; 167296B5A2E8C476 CRC64;

Query Match 65.5%; Score 78; DB 1; Length 1365;  
Best Local Similarity 71.4%; Pred. No. 0.00044;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 ANDHLSILEMSDNDPPLYLD 21  
DB 467 AIDHLSILEMSGNDNDVYKD 487

RESULT 6  
GTFD\_STRMU STANDARD; PRT; 1462 AA.  
ID GTFD\_STRMU  
AC P49331; O69383; O69386; O69389; O69392; O69398;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
GN GTFD.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-GS-5;  
RX MEDLINE=91100958; PubMed=2148600;  
RA Honda O., Kato C., Kuramitsu H.K.;  
RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding  
the glucosyltransferase-S enzyme.";  
RL J. gen. Microbiol. 136:2099-2105(1990).  
CC [2]  
CC SEQUENCE FROM N.A.  
CC STRAIN-MT4239, MT4245, MT4251, MT4467, AND MT8148;  
CC MEDLINE=98231643; PubMed=9570124;  
CC Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,  
CC Kimura S., Hamada S.;  
CC "Molecular analyses of glucosyltransferase genes among strains of  
Streptococcus mutans.";  
CC FEMS Microbiol. Lett. 161:331-336(1998).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =  
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- DISEASE: DENTAL CARIES.  
CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA  
1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES  
WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH  
FORMS OF GLUCANS.  
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
BINDING PROTEIN FROM S. MUTANS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----



```

CC EMBL: M29296; AAA26895.1; -
DR EMBL: D88653; BAA26103.1; -
DR EMBL: D88656; BAA26107.1; -
DR EMBL: D88659; BAA26111.1; -
DR EMBL: D88662; BAA26115.1; -
DR EMBL: D89979; BAA26121.1; -
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 11.
DR Pfam: PF0324; Glyco_hydro_70; 1.
KM Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL; 1
FT CHAIN ? 1462
FT DOMAIN 1232 1423 3 X 63 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 1232 1295 1.
FT REPEAT 1296 1359 2.
FT REPEAT 1360 1423 3.
FT VARIANT 58 58 K -> E (IN STRAIN MT4467).
FT VARIANT 68 68 A -> S (IN STRAINS MT4239 AND MT4245).
FT VARIANT 81 81 A -> T (IN STRAINS MT4251 AND MT8148).
FT VARIANT 113 113 T -> I (IN STRAINS MT4239 AND MT4245).
FT VARIANT 122 122 A -> V (IN STRAINS MT4239, MT4245 AND MT8148).
FT VARIANT 132 132 S -> A (IN STRAINS MT4239, MT4245, MT4251 AND MT8148).
FT VARIANT 135 135 A -> V (IN STRAIN MT4245).
FT VARIANT 202 202 V -> L (IN STRAIN MT4239).
FT VARIANT 235 235 D -> N (IN STRAIN MT8148).
FT VARIANT 275 275 E -> D (IN STRAINS MT4239, MT4245 AND MT4251).
FT VARIANT 288 288 D -> N (IN STRAINS MT4239, MT4245 AND MT4251).
FT VARIANT 301 301 Q -> H (IN STRAIN MT4245).
FT VARIANT 313 313 D -> N (IN STRAINS MT4239 AND MT4251).
FT VARIANT 317 317 E -> K (IN STRAIN MT4239).
FT VARIANT 328 328 V -> F (IN STRAIN MT4239).
FT VARIANT 350 350 F -> L (IN STRAINS MT4239, MT4251 AND MT4467).
FT VARIANT 638 638 KKKYTO -> EKEVTL (IN STRAIN MT4251).
FT VARIANT 688 688 A -> S (IN STRAIN MT4239).
FT VARIANT 726 726 TDGSGEA -> ADKNGDS (IN STRAIN MT4251).
FT VARIANT 730 730 TDGGS -> ADKGN (IN STRAINS MT4239 AND MT4245).
FT VARIANT 964 964 E -> Y (IN STRAIN MT4251).
FT VARIANT 1019 1019 E -> K (IN STRAINS MT4245 AND MT4251).
FT VARIANT 1059 1059 LG -> IR (IN STRAIN MT4251).
FT VARIANT 1060 1060 G -> R (IN STRAIN MT4245).
FT VARIANT 1080 1080 G -> R (IN STRAIN MT4239).
FT VARIANT 1142 1142 Q -> H (IN STRAINS MT4239, MT4245, MT4251, MT4467 AND MT8148).
FT VARIANT 1198 1198 S -> N (IN STRAIN MT4239).
FT VARIANT 1220 1220 Y -> C (IN STRAINS MT4251 AND MT4467).
FT VARIANT 1280 1280 F -> L (IN STRAIN MT4467).
FT VARIANT 1282 1282 K -> P (IN STRAIN MT4245).
FT VARIANT 1290 1290 Q -> T (IN STRAIN MT4245).
FT VARIANT 1311 1311 N -> G (IN STRAIN MT4245).
FT VARIANT 1403 1403 D -> G (IN STRAINS MT4239, MT4245, MT4251 AND MT8148).
FT VARIANT 1425 1425 R -> G (IN STRAINS MT4239, MT4245, MT4251, MT4467 AND MT8148).
FT VARIANT 1449 1449 R -> K (IN STRAIN MT4467).
FT CONFLICT 1448 1462 RYDKNSGMVYKVVTLANGRIIDRGIARIY -> VTR (IN REF. 1).
SQ SEQUENCE 1462 AA; 163512 MW; 5C6541F0DCB0DF00 CRC64;

```

```

Query Match 63.0%; Score 75; DB 1; Length 1462;
Best Local Similarity 71.4%; Pred. No. 0.0014;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
OY 1 ANDHLSTLEAWSNDNTPYLDH 21
I :||||| 1

```

```

DB 495 AINHLSTLEAWSNDNPOYNKD 515
RESULT 7
ID G3PC_LEIME STANDARD; PRT; 330 AA.
AC 001558;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, CYTOSOLIC (EC 1.2.1.12)
DE (GAPDH).
GN GAPC.
OS Leishmania mexicana.
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5665;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=SSP. MEXICANA;
RX MEDLINE=93063042; PubMed=1435864;
RA Hannaert V., Blaauw M., Kohl L., Allert S., Opperdoes F.R.,
RA Michaels P.A.M.;
RT Molecular analysis of the cytosolic and glycosomal glyceraldehyde-3-
RT phosphate dehydrogenase in Leishmania mexicana.
Mol. Biochem. Parasitol. 55:115-126(1992).
RL -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE
+ NAD(+) = 1,3-BISPHOSPHATEGLYCERATE + NADH.
CC -1- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
DEHYDROGENASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X65220; CAA46323.1; -
DR PIR: S25142; S25142.
DR PIR: B48445; B48445.
DR HSSP: P06977; 1GAD.
DR InterPro: IPR000173; GAP_DH.
DR Pfam: PF00044; gpdh; 1.
DR PRINTS: PR00078; G3PDHDEGNASE.
DR PROSITE: PS00071; GAPDH; 1.
KW Glycolysis; Oxidoreductase; NAD.
FT INT_MET 0
FT BINDING 148 148 BY SIMILARITY.
FT ACT_SITE 175 175 GLYCERALDEHYDE 3-PHOSPHATE.
FT SEQUENCE 330 AA; 35511 MW; EDAB6D8E8A207F1E CRC64;

```

```

Query Match 47.9%; Score 57; DB 1; Length 330;
Best Local Similarity 52.6%; Pred. No. 0.15;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
OY 2 NDHLSTLEAWSNDNTPYLDH 20
DB 300 NDHFKLVSWYDNETGYSH 318

```

```

RESULT 8
ID UL16_EBV STANDARD; PRT; 336 AA.
AC P03221;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE BGLF2 PROTEIN.

```



```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation on its
CC The European Bioinformatics Institute. There are no restrictions on files
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X81099; CAA58155.1; -.
DR SGO; Z72805; CAA97130.1; -.
DR SGD; S0003352; SEC35.
KW Hypothetical protein.
SQ SEQUENCE 275 AA; 31799 MW; AA102D086F3FAD7 CRC64;
-----
OY Query Match 38.7%; Score 46; DB 1; Length 275;
Best Local Similarity 50.0%; Pred. No. 6.1;
Matches 7; Conservative 6; Mismatches 1; Indels 0; Gaps 0;
7 ILEAMSDNDPYIHL 20
|||:::||:::||:|
217 ILESCADSNSPIYL 230
-----
RESULT 11
G3P-CHLPN STANDARD: PRT: 335 AA.
AC G9Z7TQ; Q9JQH7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, last annotation update)
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.12) (GAPDH).
GN GAP OR GAPA OR CPN0624 OR CP0123.
OC Chlamydia pneumoniae [Chlamydiophila pneumoniae].
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydiophila.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher M., Neilson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Morn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RL from Japan and CWL029 from USA.";
NL Nucleic Acids Res. 28:2311-2314(2000).
CC -I- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ONTOPHOSPHATE
CC + NAD(+)=1,3-DIPHOSPHATGLYCERATE + NADH.
CC -I- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -I- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -I- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
CC DEHYDROGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation on its
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001647; AAD18753.1; -
DR EMBL; AE002173; AAF38006.1; -
DR EMBL; AP002547; BAA98831.1; -
DR HSSP; P06977; IGAE.
DR TIGR; CP0123; -.
DR InterPro: IPR000173; GAP_DH.
DR Pfam; PF00044; gpDh_1.
DR PRINTS; PR00078; G3PDHGRGNASE.
DR PROSITE; PS00071; GAPDH; 1.
KW Glycylalys; Oxidoreductase; NAD; Complete proteome.
FT BINDING 151 151 GLYCERALDEHYDE 3-PHOSPHATE (BY
FT ACT_SITE 178 178 SIMILARITY)
FT FT ACTIVATES THIOL GROUP DURING CATALYSIS
FT FT (BY SIMILARITY).
SQ SEQUENCE 335 AA; 36837 MW; C86DDE3AD3ADF8FE CRC64;
Query Match 38.7%; Score 46; DB 1; Length 335;
Best Local Similarity 52.9%; Pred. NO. 7.7;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
QY 2 NDHLSTLEAMSDNRPY 18
Db 303 NDREFKLYAWYDNERTGY 319
RESULT 12
PCOA_PSEAE STANDARD: PRT; 476 AA.
AC Q51508;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SALICYLATE BIOSYNTHESIS ISOCHORISMATE SYNTHASE (EC 5.4.99.6).
CN PCOA OR PA4231.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=96086939; PubMed=7500944;
RA Serino L., Reimann C., Baur H., Beyeler M., Vlaso P., Haas D.;
RT "Structural genes for salicylate biosynthesis from chorismate in
RL Pseudomonas aeruginosa.";
RN Mol. Gen. Genet. 249:217-228(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Ervin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
RA Gabber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.A., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RT opportunistic pathogen.";
CC Nature 406:959-964(2000).
CC -I- FUNCTION: INVOLVED IN THE CONVERSION OF CHORISMATE TO SALICYLATE
CC (PROBABLY).
CC -I- CATALYTIC ACTIVITY: CHORISMATE = ISOCHORISMATE.
CC -I- PATHWAY: SALICYLATE BIOSYNTHESIS.
CC -I- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO
CC TYPE AND PA4B.
```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X82644; CA57969.1; -
CC EMBL: AE004840; AAC07619.1; -
CC InterPro: IPR000350; ChoriMate_bind.
CC Pfam: PF00425; ChoriMate_bind.1.
CC ProDom: PD000779; ChoriMate_bind.1.
CC Isomerase: Complete proteome.
CC SEQUENCE 476 AA; 52071 MW; D0F6AFD9C0CF9CF5 CRC64;
OY 3 DHSLEAMSNDPYLH 20
Db 44 DPLQVFGAMDROPCLY 61

RESULT 13
GAMT_HUMAN STANDARD; PRT; 236 AA.
AC 014353;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GUANIDINOCETATE N-METHYLTRANSFERASE (EC 2.1.1.2).
GN GAMT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=96138544; PubMed=8547310;
RA Isbrandt D., von Figura K.;
RA "Cloning and sequence analysis of human guanidinocetate N-
RA methyltransferase cDNA.";
RA Biochim. Biophys. Acta 1264:265-267(1995).
RP SEQUENCE FROM N.A.
RX MEDLINE=97472276; PubMed=9325156;
RA Jene D.E., Olsen A.S., Zimmer M.;
RA "The human guanidinocetate methyltransferase (GAMT) gene maps to a
RA syntenic region on 19p13.3, homologous to band C of mouse chromosome
RA 10, but GAMT is not mutated in jittery mice.";
RA Biochem. Biophys. Res. Commun. 238:723-727(1997).
RN [3]
RP SEQUENCE FROM N.A.
RA Isbrandt D., Schmidt A.;
RA "Gene structure of human guanidinocetate N-methyltransferase.";
RA submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA Dangann L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Krommiller B., Arellano A., Montgomery K., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.;
RA submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + GUANIDOCETATE =

```

```

CC S-ADENOSYL-L-HOMOCYSTEINE + CREATINE.
CC -1- PATHWAY: LAST STEP OF CREATINE BIOSYNTHESIS.
CC DISEASE: DEFECTS IN GAMT ARE RESPONSIBLE FOR NEUROLOGIC SYNDROMES
CC AND MUSCULAR HYPOTONIA. BIOCHEMICALLY IT IS CHARACTERIZED BY LOW
CC EXCRETION OF CREATINE. DEFICIENCY OF CREATINE AND CREATINE
CC PHOSPHATE, AND SIMULTANEOUS ACCUMULATION OF GUANIDINOCETATE IN
CC BRAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z49878; CAA90035.1; -
CC EMBL: AF010248; AAD04781.1; -
CC EMBL: AF010246; AAD04781.1; JOINED.
CC EMBL: AF010247; AAD04781.1; JOINED.
CC EMBL: AF188893; AAF01461.1; -
CC EMBL: AC005329; AAC27668.1; -
CC MIM: 601240; -
CC Transferase: Methyltransferase.
CC SEQUENCE 236 AA; 26318 MW; 688E845CE56189F5 CRC64;
OY 1 ANDHSLT-----EAWSDNDPYLH 20
Db 30 ADTHRLRLGKPYMERW---ETPYMH 51

RESULT 14
GNF_FUSOX STANDARD; PRT; 385 AA.
AC P46239;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PUTATIVE ENDODGLUCANASE TYPE F PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-
DE GLUCANASE) (CELLULOSE).
OS Fusarium oxysporum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; mitosporic Hypocreales; Fusarium.
OX NCBI_TaxID=5507;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95047531; PubMed=7959045;
RA Sheppard P.O., Grant F.J., Oort P.J., Sprecher C.A., Foster D.C.,
RA Hagen F.S., Upshall A., McKnight G.L., O'Hara P.J.;
RA "The use of conserved cellulase family-specific sequences to clone
RA cellulase homologue cDNAs from Fusarium oxysporum.";
RA Gene 150:163-167(1994).
RN [2]
RP CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
RP LINKAGES IN CELLULOSE.
RP -1- SIMILARITY: CONTRAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
RP -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY F (FAMILY 10 OF GLYCOSYL
RP HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U29380; AAA65588.1; -
CC HSSP: P56588; 1BG4.

```

```

DR InterPro: IPR000254; CBD_fungal.
DR InterPro: IPR001000; Glyco_hydro_10.
DR Pfam: PF00734; CBD_1; 1.
DR Pfam: PF00331; Glyco_hydro_10; 1.
DR PRINTS: PR00134; GLHYDRLASE10.
DR PRODOM: PD001821; CBD_fungal; 1.
DR SMART: SM00236; fcbd; 1.
DR PROSITE: PS00562; CBD_FUNGAL; 1.
DR PROSITE: PS00591; GLYCOSYL_HYDROL_F10; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; Signal.
KW SIGNAL
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 385 PUTATIVE ENDOSGLUCANASE TYPE F.
FT DOMAIN 25 52 CELLULOSE-BINDING.
FT DOMAIN 53 84 LINKER.
FT DOMAIN 85 385 CATALYTIC.
FT ACT_SITE 209 209 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 321 321 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 385 AA; 41225 MW; B3C3807C07D3C0EC CRC64;

OY 1 ANDHLSIEA-----MSDNDFPYLHD 21
St Local Similarity 37.8%; Score 45; DB 1; Length 385;
Matches 10; Conservative 5; Mismatches 6; Indels 20; Gaps 1;

Db 329 ANDYITVNAACLAAPKCYGITYWGVSDKDSMRPGDNPILYD 369

RESULT 15
AMP2_STRCO STANDARD; PRT; 470 AA.
AC Q60394;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE XAA-PRO AMINOPEPTIDASE II (EC 3.4.11.9) (X-PRO AMINOPEPTIDASE II)
DE (AMINOPEPTIDASE P II) (APP) (PEPP II) (AMINOACTEPROLINE AMINOPEPTIDASE
DE II).
GN PEPP2 OR SCG61.34C.
OS Streptomyces coelicolor, and Streptomyces lividans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902, 1916;
[1]
RN SEQUENCE FROM N.A.
RC SPECIES=S.coelicolor; STRAIN=A3(2);
RA Oliver K., Harris D., Bentley S.D., Parkhill J., Barrell B.G.,
RA Rajandream M.A.;
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.lividans; STRAIN=66;
RC MEDLINE=941176084; PubMed=7765336;
RA Butler M.J., Aphale J.S., Dizonno M.A., Krygman P., Walczyk E.,
RA Malek L.T.;
RT "Intracellular aminopeptidases in Streptomyces lividans 66.";
RL J. Ind. Microbiol. 13:24-29(1994).
CC -I- CATALYTIC ACTIVITY: RELEASE OF ANY N-TERMINAL AMINO ACID,
CC INCLUDING PROLINE, THAT IS LINKED WITH PROLINE, EVEN FROM A
CC DIPEPTIDE OR TRIPEPTIDE.
CC -I- COFACTOR: ZINC OR MANGANESE.
CC -I- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -I- MISCELLANEOUS: IN STREPTOMYCES LIVIDANS, THERE ARE TWO GENES
CC (PEPP1 AND PEPP2) WHICH ENCODE AMINOPEPTIDASE P.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL359949; CAB95809.1; -.
DR EMBL: L23174; AAB00325.1; -.
DR MEROPS: M24.025; -.
DR InterPro: IPR000994; Peptidase_M24.
DR InterPro: IPR001131; Pro_dipeptidase.
DR Pfam: PF00557; Peptidase_M24; 1.
DR PROSITE: PS00491; PROLINE_PEPTIDASE; 1.
DR Hydrolyase; Aminopeptidase; Manganese; Zinc.
SQ SEQUENCE 470 AA; 51924 MW; 0D3D33CE256523C2 CRC64;

OY 1 ANDHLSIEAMSDNDTP 17
Db 262 AGEHATIMH-WTDNDGP 277

Query Match 37.4%; Score 44.5; DB 1; Length 470;
Best Local Similarity 47.1%; Pred. No. 20;
Matches 8; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

```

Search completed: March 27, 2002, 14:27:00  
Job time: 1642 sec

...

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2002, 13:58:08 ; Search time 188.53 seconds  
(without alignments)  
16.293 Million cell updates/sec

Title: US-09-290-049a-1  
Perfect score: 119  
Sequence: 1 ANDHSLILEAMSDNDPPLYLHD 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	100.0	1390	2	069385 streptococc
2	119	100.0	1455	2	069391 streptococc
3	110	92.4	1435	2	069382 streptococc
4	110	92.4	1455	2	069388 streptococc
5	110	92.4	1455	2	069397 streptococc
6	108	90.8	1590	2	055263 streptococc
7	108	90.8	1590	2	059983 streptococc
8	78	65.5	1338	2	09WJ4 streptococc
9	76	63.9	1016	2	09LJ7 leuconostoc
10	74	62.2	1477	2	09L46 leuconostoc
11	74	62.2	1508	2	052224 leuconostoc
12	74	62.2	1508	2	09EZH5 leuconostoc
13	73	61.3	1575	2	09LCH3 streptococc
14	73	61.3	1577	2	054178 streptococc
15	71	59.7	1527	2	09ZAR4 leuconostoc
16	66	55.5	1449	2	068542 streptococc
17	66	55.5	1449	2	055264 streptococc
18	63	52.9	1577	2	055265 streptococc
19	62	52.1	1512	2	09WJ5 streptococc

20	60	50.4	2057	2	09RE05 leuconostoc
21	58	48.7	1518	2	000600 streptococc
22	57	47.9	1599	2	000599 streptococc
23	52	43.7	4848	2	007944 streptomyc
24	51	42.9	435	10	09L068 arabidopsis
25	50	42.0	347	10	09LY92 arabidopsis
26	49.5	41.6	549	10	043274 zea mays (m
27	49	41.2	601	2	005566 mycobacteri
28	48	40.3	490	2	006579 mycobacteri
29	47.5	39.9	519	10	09Z0B6 arabidopsis
30	47.5	39.9	549	10	09L0R2 oryza sativ
31	47.5	39.9	549	10	09FRX7 oryza sativ
32	47	39.5	175	10	09LM08 arabidopsis
33	47	39.5	378	10	09SZH3 arabidopsis
34	47	39.5	524	2	09KST8 vibrio chol
35	47	39.5	525	3	059679 schizosacch
36	46	38.7	195	13	09DFN6 gillilichys
37	46	38.7	2567	5	09U0U6 plasmodium
38	45.5	38.2	96	2	09JST4 neisseria m
39	45	37.8	286	2	09RDG2 streptomyc
40	45	37.8	333	13	09PTW5 arabidopsi
41	45	37.8	947	10	023136 arabidopsi
42	45	37.8	1159	4	09UFC3 homo sapien
43	44.5	37.4	210	2	09I0Y4 pseudomonas
44	44.5	37.4	499	5	09VJ56 drosophila
45	44	37.0	44	4	016768 homo sapien

#### ALIGNMENTS

RESULT 1  
ID 069385 PRELIMINARY: PRT; 1390 AA.  
AC 069385;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTFC.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MT4245;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S., Kimura S., Hamada S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans.";  
RL FEMS Microbiol. Lett. 161:331-336(1998).  
DR EMBL: D88655; BAA26106.1; .  
DR InterPro: IPR002479; CW\_binding.  
DR InterPro: IPR003318; Glyco\_hydro\_70.  
DR Pfam: PF01473; CW\_binding\_1; 7.  
DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
KW Transferase.  
SO SEQUENCE 1390 AA; 155375 MW; 8847E4956EF05E9F CRC64;

Query Match 100.0%; Score 119; DB 2; Length 1390;  
Best Local Similarity 100.0%; Pred. No. 7.5e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ANDHSLILEAMSDNDPPLYLHD 21  
DB 507 ANDHSLILEAMSDNDPPLYLHD 527  
RESULT 2  
069391

ID 069391 PRELIMINARY; PRT: 1455 AA.  
 AC 069391;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-ST.  
 GN GTF.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MT4251;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,  
 Kimura S., Hamada S.;  
 "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans.";  
 FEMS Microbiol. Lett. 161:331-336(1998).  
 DR EMBL: D88661; BAA26114.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 9.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 SQ SEQUENCE 1455 AA; 162804 MW; 683A359D873E9E1A CRC64;

Query Match 100.0%; Score 119; DB 2; Length 1455;  
 Best Local Similarity 100.0%; Pred. No. 7.9e-10;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 ANDHLSILEAMSDNDTPYLHD 21  
 DB 507 ANDHLSILEAMSDNDTPYLHD 527

RESULT 3  
 ID 069382 PRELIMINARY; PRT: 1455 AA.  
 AC 069382;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-ST.  
 GN GTF.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MT8148;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,  
 Kimura S., Hamada S.;  
 "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans.";  
 FEMS Microbiol. Lett. 161:331-336(1998).  
 DR EMBL: D88652; BAA26102.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 9.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 SQ SEQUENCE 1455 AA; 162969 MW; 27DAD3A1ECCA2939 CRC64;

Query Match 92.4%; Score 110; DB 2; Length 1455;  
 Best Local Similarity 95.2%; Pred. No. 2e-08;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ANDHLSILEAMSDNDTPYLHD 21  
 DB 507 ANDHLSILEAMSYNDTPYLHD 527

RESULT 4  
 ID 069388 PRELIMINARY; PRT: 1455 AA.  
 AC 069388;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-ST.  
 GN GTF.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MT4239;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,  
 Kimura S., Hamada S.;  
 "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans.";  
 FEMS Microbiol. Lett. 161:331-336(1998).  
 DR EMBL: D88658; BAA26110.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 10.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 SQ SEQUENCE 1455 AA; 163045 MW; 6D90A4978D35DD82 CRC64;

Query Match 92.4%; Score 110; DB 2; Length 1455;  
 Best Local Similarity 95.2%; Pred. No. 2e-08;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ANDHLSILEAMSDNDTPYLHD 21  
 DB 507 ANDHLSILEAMSYNDTPYLHD 527

RESULT 5  
 ID 069397 PRELIMINARY; PRT: 1455 AA.  
 AC 069397;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE-ST.  
 GN GTF.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MT4467;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,  
 Kimura S., Hamada S.;  
 "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans.";  
 FEMS Microbiol. Lett. 161:331-336(1998).  
 DR EMBL: D89978; BAA26120.1; -  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR003318; Glyco\_hydro\_70.  
 DR Pfam: PF01473; CW\_binding\_1; 9.  
 DR Pfam: PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.



SO SEQUENCE 1455 AA; 162913 MW; A1263427BF24E8E1 CRC64;

Query Match 92.4%; Score 110; DB 2; Length 1455;

Best Local Similarity 95.2%; Pred. No. 2e-08;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ANDHSTILEAWSNDNDTPYLHD 21

DB 507 ANDHSTILEAWSNDNDTPYLHD 527

RESULT 6

ID 055263 PRELIMINARY; PRT: 1590 AA.

AC 035263;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE GTF-1.

Streptococcus sobrinus.

Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

Streptococcus.

NCBI\_TaxID=1310;

OX 11

RN SEQUENCE FROM N.A.

RC STRAIN=ATCC 33478;

RA Sato S.;

RL Ann. Kagoshima Univ. Dental School 16:23-29(1996).

DR EMBL; D63570; BAA09792.1; "

DR Interpro: IPR002479; CW\_binding.

DR Interpro: IPR003318; Glyco\_hydro\_70.

DR Pfam; PF01473; CW\_binding\_1; 15.

DR Pfam; PF02324; Glyco\_hydro\_70; 1.

KM Transferase.

SO SEQUENCE 1590 AA; 176057 MW; 9DF7A3F2C6E4FD43 CRC64;

Query Match 90.8%; Score 108; DB 2; Length 1590;

Best Local Similarity 85.7%; Pred. No. 4.5e-08;

Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHSTILEAWSNDNDTPYLHD 21

DB 477 ANNHVSIVEAWSNDNDTPYLHD 497

RESULT 7

ID 059983 PRELIMINARY; PRT: 1590 AA.

AC 059983;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE

6-GLUCOSYLTRANSFERASE).

GN GTF1.

OS Streptococcus sobrinus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1310;

RN SEQUENCE FROM N.A.

RC STRAIN=OM2176;

RX MEDLINE=94146405; PubMed=8312602;

RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;

RT "DNA sequence of the glucosyltransferase gene of serotype d

Streptococcus sobrinus.";

RL DNA Seq. 4:19-27(1993).

-i- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-

FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).

EMBL; D13858; BAA02976.1; "

DR Interpro: IPR002479; CW\_binding.

DR Interpro: IPR003318; Glyco\_hydro\_70.

DR Pfam; PF01473; CW\_binding\_1; 16.

DR Pfam; PF02324; Glyco\_hydro\_70; 1.

KW Signal; Transferase; Glycosyltransferase.

FT SIGNAL 1 38 POTENTIAL.

CHAIN 39 1590 GLUCOSYLTRANSFERASE-1.

SO SEQUENCE 1590 AA; 175955 MW; C3C83A57C3C2B0E CRC64;

Query Match 90.8%; Score 108; DB 2; Length 1590;

Best Local Similarity 85.7%; Pred. No. 4.5e-08;

Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 ANDHSTILEAWSNDNDTPYLHD 21

DB 477 ANNHVSIVEAWSNDNDTPYLHD 497

RESULT 8

ID 09MXJ4 PRELIMINARY; PRT: 1338 AA.

AC 09MXJ4;

DT 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE GTF-S.

GN Streptococcus criceti.

OS Plasmid pAM1.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1333;

RN SEQUENCE FROM N.A.

RC STRAIN=HS-6;

RA Inoue M., Fukui K., Miyagi A.;

RT "S. cricetus glucosyltransferase(gtfs and gtf) genes.";

RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.

DR EMBL; AB026123; BAA77236.1; "

DR HSSP; P06278; IVS.

DR Interpro: IPR002479; CW\_binding.

DR Interpro: IPR003318; Glyco\_hydro\_70.

DR Pfam; PF01473; CW\_binding\_1; 10.

DR Pfam; PF02324; Glyco\_hydro\_70; 1.

KM Plasmid.

SO SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;

Query Match 65.5%; Score 78; DB 2; Length 1338;

Best Local Similarity 71.4%; Pred. No. 0.0017;

Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 ANDHSTILEAWSNDNDTPYLHD 21

DB 437 AIDHSTILEAWSNDNDYKVD 457

RESULT 9

ID 09LCJ7 PRELIMINARY; PRT: 1016 AA.

AC 09LCJ7;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE DEXTRANSUCRASE.

GN DSR1.

OS Leuconostoc mesenteroides.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;

OC Leuconostoc.

OX NCBI\_TaxID=1245;

RN SEQUENCE FROM N.A.

RC STRAIN=NRRL B-512F;

RX MEDLINE-20169623; PubMed-10705445;  
 RA Funane K., Mizuno K., Takahara H., Kobayashi M.;  
 RT "Gene encoding a dextranucrase-like protein in Leuconostoc  
 RL mesenteroides NRRL B-512F."; 64:29-38(2000).  
 DR EMBL; AB020020; BAA90527.1; -;  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1016 AA; 110343 MW; 8896EPDE13CCBA7 CRC64;

Query Match 63.9%; Score 76; DB 2; Length 1016;  
 Best Local Similarity 61.9%; Pred. No. 0.0025;  
 Matches 13; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 ANDHLSLEAMSDNDPTYLHD 21  
 :||:|||||:|||||:|:  
 548 ANQHLSTLEDMSHNDPEYKVD 568

RESULT 10

ID 09L466 PRELIMINARY; PRT; 1477 AA.

AC 09L466; 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE (EC 2.4.1.5).  
 GN DSRG.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NRRL B-1355;  
 RA Arguello-Morales M.A., Remaud-Simeon M., Plizut S., Sarcabal P.,  
 RA Willemot R.M., Monsan P.;  
 RT "Sequence analysis of the gene encoding alternansucrase, a sucrose  
 RT glucosyltransferase from Leuconostoc mesenteroides NRRL B-1355.";  
 RL Submitted (OCF-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ250172; CAB76565.1; -;  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 14.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR Transferase; Glycosyltransferase.  
 SQ SEQUENCE 1477 AA; 164886 MW; E6F5710DEDFC8B31 CRC64;

Query Match 62.2%; Score 74; DB 2; Length 1477;  
 Best Local Similarity 66.7%; Pred. No. 0.0079;  
 Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 ANDHLSLEAMSDNDPTYLHD 21  
 |||||||:|||||:|:  
 DB 532 ANQHLSTLEDMSHNDPEYKVD 552

RESULT 11

ID 052224 PRELIMINARY; PRT; 1508 AA.

AC 052224; 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE 6-  
 DE GLUCOSYLTRANSFERASE).  
 GN DSRB.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.  
 OX NCBI\_TaxID=1245;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NRRL B-1299;  
 RA Monchois V., Remaud-Simeon M., Monsan P., Willemot R.M.;  
 RL FEMS Microbiol. Lett. 0:0-0(1998).  
 CC -I- CATALYTIC ACTIVITY: SUCROSE + ((1,6)-ALPHA-D-GLUCOSYL)(N) = D-  
 CC FRUCTOSE + ((1,6)-ALPHA-D-GLUCOSYL)(N+1).  
 DR EMBL; AF030129; AAB95453.1; -;  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 14.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR Transferase; Glycosyltransferase.  
 SQ SEQUENCE 1508 AA; 168511 MW; E70CECB57A70D1F0 CRC64;

Query Match 62.2%; Score 74; DB 2; Length 1508;  
 Best Local Similarity 66.7%; Pred. No. 0.0081;  
 Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 ANDHLSLEAMSDNDPTYLHD 21  
 |||||||:|||||:|:  
 DB 563 ANQHLSTLEDMSHNDPEYKVD 583

RESULT 12

ID 09EZHS PRELIMINARY; PRT; 1508 AA.

AC 09EZHS; 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE DSRB742.  
 GN DSRB742.  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-B-742CB;  
 RA Kim H.-S., Kim D., Ryu H.-J., Robyt J.F.;  
 RT "Leuconostoc mesenteroides B-742CB, a dextranucrase gene."  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF294469; AAG38021.1; -;  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 14.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1508 AA; 168542 MW; E2FCFA0F87AE4F3A CRC64;

Query Match 62.2%; Score 74; DB 2; Length 1508;  
 Best Local Similarity 66.7%; Pred. No. 0.0081;  
 Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 ANDHLSLEAMSDNDPTYLHD 21  
 |||||||:|||||:|:  
 DB 563 ANQHLSTLEDMSHNDPEYKVD 583

RESULT 13

ID 09LCH3 PRELIMINARY; PRT; 1575 AA.

AC 09LCH3; 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE.  
 GN GTPP.  
 OS Streptococcus oralis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.

OX NCBI\_TaxID=1303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC10537;  
 RA MEDLINE=20231779; PubMed=10768934;  
 RX Fujiwara T., Hoshino T., Oshima T., Sobue S., Hamada S.;  
 RT "Purification, characterization, and molecular analysis of the gene  
 encoding glucosyltransferase from *Streptococcus oralis*.";  
 RL Infect Immun 68:2475-2483(2000).  
 DR EMBL; AB025228; BAA95201.1; -;  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 17.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR Transferrase;  
 KW  
 SQ SEQUENCE 1575 AA; 176792 MM; 772A26E4D/C2E543 CRC64;

Query Match 61.3%; Score 73; DB 2; Length 1575;  
 Best Local Similarity 77.8%; Pred. No. 0.012;  
 Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 4 HLSILEAWSNDNDPYLHD 21  
 |||||  
 Db 549 HLSILEAWSNDNDPYNKD 566

RESULT 14  
 ID 054178 PRELIMINARY; PRT; 1577 AA.  
 AC 054178; 054247;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE GLUCOSYLTRANSFERASE.  
 GN GTFG.  
 OS *Streptococcus gordonii* Challis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 CC Streptococcus  
 OX NCBI\_TaxID=29390;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CHALLIS;  
 RX MEDLINE=96157084; PubMed=8586195;  
 RA Vickerman M.M., Sulavik M.C., Clewell D.B.;  
 RT "Molecular analysis of *Streptococcus gordonii* glucosyltransferase  
 phase variants.";  
 RL Dev. Biol. Stand. 85:309-314(1995).  
 RN [2]  
 RP SEQUENCE OF 1-96 FROM N.A.  
 RC STRAIN=CHALLIS;  
 RX MEDLINE=92276337; PubMed=1534326;  
 RA Sulavik M.C., Tardif G., Clewell D.B.;  
 RT "Identification of a gene, rgg, which regulates expression of  
 glucosyltransferase and influences the Spp phenotype of *Streptococcus*  
*gordonii* Challis.";  
 RL J. Bacteriol. 174:3577-3586(1992).  
 DR EMBL; U12643; AAC43483.1; -;  
 DR EMBL; M89776; AAA26969.1; -;  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 18.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR Transferrase;  
 KW  
 SQ SEQUENCE 1577 AA; 177805 MM; 5AE032BDC5E08D18 CRC64;

Query Match 61.3%; Score 73; DB 2; Length 1577;  
 Best Local Similarity 77.8%; Pred. No. 0.012;  
 Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 4 HLSILEAWSNDNDPYLHD 21  
 |||||

Db 551 HLSILEAWSNDNDPYNKD 568

RESULT 15  
 ID 092AR4 PRELIMINARY; PRT; 1527 AA.  
 AC 092AR4;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE DEXTRANSUCRASE.  
 GN DEX.  
 OS *Leuconostoc mesenteroides*.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 CC Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL B-512-F;  
 RA Bhatnagar R., Singh D.K.S.;  
 RT "Cloning and Molecular Characterization of Dextranucrase Gene from  
*Leuconostoc mesenteroides* NRRL B-512F.";  
 RL Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; U81374; AAD10952.1; -;  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 16.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR Transferrase;  
 KW  
 SQ SEQUENCE 1527 AA; 169709 MM; 1DFAFA237C743398 CRC64;

Query Match 59.7%; Score 71; DB 2; Length 1527;  
 Best Local Similarity 66.7%; Pred. No. 0.024;  
 Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 1 ANDHLSILEAWSNDNDPYLHD 21  
 |||||  
 Db 581 ANQHLSILEAWSNDNDPYLHD 601

Search completed: March 27, 2002, 14:26:01  
 Job time: 1673 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: March 27, 2002, 13:57:52 ; Search time 198.55 Seconds  
(without alignments)  
7.088 Million cell updates/sec

Title: US-09-290-049a-2  
Perfect score: 95  
Sequence: 1 VPSYSTRRAHSEVODLIA 19

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Number of hits satisfying chosen parameters: 522463

Maximum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_1101.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT:\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT:\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	84	88.4	1592	14	AA32925	Glucosyltransferase
2	66	69.5	2057	21	AA31067	L. mesenteroides a
3	61	64.2	1577	17	AA391047	Alpha-D-glucosyltr
4	42	44.2	616	21	AA32861	Haemophilus influe
5	42	44.2	2625	19	AA55887	Human telomerase
6	42	44.2	2627	19	AA61347	Human telomerase R
7	41.5	43.7	713	17	AA39797	Lysine decarboxyla
8	41	43.2	170	22	AA30785	Human dual-specific
9	41	43.2	344	14	AA40918	Bacillus subtilis
10	41	43.2	552	22	AA30772	Human dual-specific
11	41	43.2	552	22	AA307043	Human dual-specific

12	41	43.2	662	22	AA32736	C glutamicum prote
13	41	43.2	1049	22	AA304833	Human SCP006 phosph
14	41	43.2	1120	20	AA319780	B. burgdorferi ant
15	41	43.2	1123	21	AA358277	Helicobacter armiger
16	41	43.2	1146	20	AA319779	B. burgdorferi ant
17	40	42.1	149	21	AA340856	Human ORF620
18	40	42.1	170	22	AA306786	Human dual-specific
19	40	42.1	241	22	AA306776	Human dual-specific
20	40	42.1	359	22	AA306263	RetA protein of CT
21	40	42.1	431	11	AA307464	Polypeptide with e
22	40	42.1	431	20	AA307890	Protein encoded by
23	40	42.1	484	22	AA385363	Human phosphatase
24	40	42.1	498	22	AA304835	Human SCP001 phosph
25	40	42.1	509	22	AA306775	Human dual-specific
26	40	42.1	509	22	AA307045	Human dual-specific
27	40	42.1	650	22	AA38706	Human polypeptide
28	40	42.1	759	22	AA340492	Human polypeptide
29	40	42.1	759	22	AA340492	Human polypeptide
30	39	41.1	174	22	AA301115	Human gene 2 encod
31	39	41.1	356	22	AA382330	S. epidermidis ope
32	39	41.1	446	20	AA325684	Cockroach allergen
33	39	41.1	489	22	AA300051	C glutamicum prote
34	39	41.1	946	22	AA3070686	S. cerevisiae apopt
35	39	41.1	2058	22	AA397070	Human polypeptide
36	38.5	40.5	237	16	AA384562	Human prostacyclin
37	38.5	40.5	310	22	AA394596	Human protein sequ
38	38.5	40.5	345	20	AA317246	Human protein kina
39	38.5	40.5	350	21	AA318887	A prolactin regula
40	38.5	40.5	417	21	AA318886	Human prolactin
41	38.5	40.5	417	21	AA394409	Human protein sequ
42	38.5	40.5	500	16	AA384561	Human prostacyclin
43	38.5	40.5	557	19	AA320666	Human neurofilamen
44	38	40.0	166	22	AA356026	Peptide #10063 enc
45	38	40.0	229	21	AA323447	Arabidopsis thalia

#### ALIGNMENTS

XX	RESULT 1	
XX	AA32925	
ID	AA32925	standard; Protein: 1592 AA.
XX		
AC	AA32925;	
XX		
DT	28-JUN-1993	(first entry)
XX		
DE	Glucosyltransferase I.	
XX		
KW	GT-I; Streptococcus; dental; caries.	
XX		
OS	Streptococcus sobrinus.	
XX		
PN	JP05023188-A.	
XX		
PD	02-FEB-1993.	
XX		
PE	25-JUL-1991;	91JP-0186592.
XX		
PR	25-JUL-1991;	91JP-0186592.
XX		
PA	(FUKU/) FUKU I.	
PA	(KATO/) KATO K.	
XX		
DR	WPI: 1993-079449/10;	
DR	N-PSDB; AAQ37760.	
XX		
PT	DNA sequence glucosyl:transferase-I - comprises Streptococcus	
PT	sobrinus DNA sequence with at least one nucleotide added or	
XX	deleted	
PS	Claim 13; Page 15; 29pp; Japanese.	
XX		

CC The DNA sequence from *Streptococcus sobrinus* strain 6715 encodes  
 CC glucosyltransferase-I (and mutants). The DNA was obtd. by treating  
 CC *S. sobrinus* 6715 with mutanolysin, extracting the chromosomal DNA,  
 CC partially digesting with Sau3AI and fractionating on agarose gel.  
 CC The 3-5 kbp fragment was ligated into pUC18 and *E. coli* JM109  
 CC transformed with it. A GT-1 expressing clone was isolated and  
 CC sequenced. The clone may be used in the development of a drug for  
 CC dental caries.  
 CC  
 SQ Sequence 1592 AA:

Query Match 88.4%; Score 84; DB 14; Length 1592;  
 Best Local Similarity 88.9%; Pred. No. 1.7e-06;  
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VPYSFIRAHDSVODLI 18  
 ||||| |||||  
 548 vpsfahsdvqdli 565

## RESULT 2

AAB10667  
 ID AAB10667 standard; Protein: 2057 AA.

AC AAB10667;

DT 19-JAN-2001 (first entry)

DE L. mesenteroides alternan sucrose protein.

KW Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;  
 KM syrup.

OS Leuconostoc mesenteroides.

PN DE19905069-A1.

PD 10-AUG-2000.

PF 08-FEB-1999; 99DE-1005069.

PR 08-FEB-1999; 99DE-1005069.

PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.

PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

Kossmann J, Welsh T, Quanz M, Knuth K;

DR WPI: 2000-550294/51.

DR N-PSDB: AAA97904.

PT New nucleic acid encoding recombinant Leuconostoc mesenteroides  
 PT alternan sucrose protein and methods of alternan and fructose  
 PT production

PS Claim 1a; Page 30-36; 64pp; German.

CC This invention describes a novel nucleic acid molecule (I) encoding an  
 CC alternan sucrose (E.C. 2.4.1.140 - an enzyme, that belongs to the  
 CC glucosyltransferase group). The recombinant, purified alternan sucrose  
 CC gene is useful for the fermentative production of alternan (a  
 CC carbohydrate) and/or fructose by secreting the enzyme into a  
 CC saccharose-containing culture medium. Alternatively, the enzyme is  
 CC contacted with a saccharose-containing solution. The alternan and/or  
 CC fructose is then isolated from the medium. Cosmetic products or  
 CC foodstuffs containing alternan can be produced. Recombinant production of  
 CC alternan sucrose is advantageous as it provides a cost effective means of  
 CC producing fructose for high fructose containing syrups, production of  
 CC which previously has been achieved by costly production from maize  
 CC starch. This sequence represents the leuconostoc mesenteroides alternan  
 CC sucrose protein which is described in the method of the invention.

SQ Sequence 2057 AA;

Query Match 69.5%; Score 66; DB 21; Length 2057;  
 Best Local Similarity 61.1%; Pred. No. 0.0049;  
 Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 VPYSFIRAHDSVODLI 18  
 :|:|:|:|:|:|:|:|  
 Db 757 ipnysfirahdydaqdpi 774

## RESULT 3

AAR91047  
 ID AAR91047 standard; Protein: 1577 AA.

AC AAR91047;

DT 22-MAY-1996 (first entry)

DE Alpha-D-glucosyltransferase.

KW Alpha-D-glucosyltransferase; primer-independent; soluble glucan;  
 KM sucrose; transgenic plant; cloning; *Escherichia coli*;

KW phage lambda-cl3; vector; plasmid pGSG501; plasmid pGSG502;

KW gene transfer; crop improvement; storage carbohydrate; pasture;  
 KM feedstuff; senescence; dextran; binder; food; pharmaceutical.

OS *Streptococcus salivarius* strain ATCC 25975.

PN W09606173-A1.

PD 29-FEB-1996.

PF 24-AUG-1995; 95WO-AU00527.

PR 24-AUG-1994; 94AU-0007643.

PA (GIFE/) GIFFARD P M.

PA (JACO/) JACOUES N A.

PA (SIMP/) SIMPSON C L.

PI Giffard PM, Jacques NA, Simpson CL;

DR WPI: 1996-151376/15.

DR N-PSDB: AAT13139.

PT plants contg. new bacterial DNA encoding glucosyl transferase  
 PT activity - retain higher levels of stored carbohydrate(s) in a form  
 PT readily digestible by ruminants

PS Claim 4; Page 16-20; 31pp; English.

CC The sequence represents an alpha-D-glucosyltransferase from  
 CC *Streptococcus salivarius*. The enzyme is primer-independent, and  
 CC produces soluble glucan from sucrose. A gene encoding the enzyme  
 CC may be cloned and expressed in *Escherichia coli* using a subclone  
 CC of phage lambda-cl3, e.g. plasmid pGSG501 or plasmid pGSG502. The  
 CC DNA may also be expressed in a transgenic plant, to improve the  
 CC level of stored carbohydrate in a pasture plant which normally  
 CC contains low levels, or to prevent degradation of stored carbohydrate  
 CC during plant senescence. Dextran may be isolated from the plant, for  
 CC use as a food binder or pharmaceutical additive. Primer independence  
 CC ensures that the enzyme will be functional in plants. The glucan is  
 CC poorly degraded in plants but easily degraded by bacteria in the rumen  
 CC of grazing livestock.

SQ Sequence 1577 AA;

Query Match 64.2%; Score 61; DB 17; Length 1577;  
 Best Local Similarity 64.7%; Pred. No. 0.03;  
 Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 3 SYSFIRAHDSFVODLIA 19  
:|:|||||||:  
Db 661 nyifvrahdsfvgavla 677

# RESULT 4

AAB23861  
ID AAB23861 standard; Protein; 616 AA.

AC AAB23861;

DT 17-JAN-2001 (first entry)

DE Haemophilus influenzae adhesin (Hia) protein from NTH1 strain 12.

KW Hia; adhesin; Haemophilus influenzae adhesin; NTH1; infection; vaccine;  
KW non-typable Haemophilus influenzae; antiinflammatory; auditory;  
KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;  
KW diagnosis; immunogenic; antigen.

Haemophilus influenzae.

WO200055191-A2.

21-SEP-2000.

16-MAR-2000; 2000WO-CA00289.

16-MAR-1999; 99US-0268347.

(CONN-) CONNAUGHT LAB LTD.

Loosmore SM, Yang Y, Klein MH;

WPI; 2000-618897/59.

N-PSDB; AAA92500.

Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for use as antigens and vaccines and for treating Hemophilus influenzae infection.

Claim 1; Fig 25; 275pp; English.

The present sequence represents a Haemophilus influenzae adhesin (Hia) protein from the non-typable Haemophilus influenzae (NTH1) strain 12. Hia genes and proteins have antiinflammatory, auditory and antibacterial activities, and can be used in the production of a vaccine. An immunogenic composition comprising an Hia gene, a polypeptide encoded by an Hia gene, or a recombinant Hia polypeptide is useful for inducing protection against disease caused by Haemophilus strains in a susceptible host, preferably a human. An Hia protein is useful as an antigen, in immunogenic preparations including vaccines, as a carrier for other immunogens, and in the generation of diagnostic reagents. Hia is useful for treating diseases caused by the infection of Haemophilus influenzae such as meningitis, epiglottitis, septicaemia and otitis media. Recombinant production of Hia favours high recovery of the protein compared to the low recovery of native protein from Haemophilus influenzae species. A truncated protein has a significantly higher amount of recovery than a full-length protein.

Sequence 616 AA;

Query Match 44.2%; Score 42; DB 21; Length 616;  
Best Local Similarity 38.9%; Pred. No. 32;  
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 2 PSYSFIRAHDSFVODLIA 19  
|:|:|:|:|:|:|:  
Db 42 polyatvlaengkvqelka 59

# RESULT 5

AAW55887  
ID AAW55887 standard; Protein; 2625 AA.

AC AAW55887;

DT 22-JUL-1998 (first entry)

DE Human telomerase.

KW Rat; telomerase; human; cancer; screening; inhibitor; elucidation;  
KW detection; probe; diagnosis; cell growth; ageing.

OS Homo sapiens.

WO9807838-A1.

26-FEB-1998.

21-AUG-1997; 97WO-JP02904.

17-FEB-1997; 97JP-0031807.

21-AUG-1996; 96JP-0219761.

31-JAN-1997; 97JP-0018878.

(MITU) MITSUBISHI CHEM CORP.

Fujino Y, Harada N, Ishikawa F, Nakamura H, Takahashi K;

WPI; 1998-169149/15.

N-PSDB; AAV25989.

Telomerase protein of higher animals and humans and gene encoding it - for use in diagnosis of cancer, screening of telomerase inhibitors and elucidation of biological control mechanisms

Claim 8; Page 74-93; 106pp; Japanese.

The present sequence represents a human protein component of telomerase. The DNA or RNA encoding the telomerase protein component or its fragments can be used as a nucleotide probe for the detection of cancer cells and for diagnosis of cancer. Potential telomerase inhibitors can be screened by measuring their effect on the assay of the active form in cells or tissues. The polypeptide and DNA coding for it can be used in the elucidation of biological control mechanisms of, e.g. cell growth or ageing and of the mechanisms of cancer development.

Sequence 2625 AA;

Query Match 44.2%; Score 42; DB 19; Length 2625;  
Best Local Similarity 47.1%; Pred. No. 1.8e+02;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 1 VPSYSFIRAHDSFVODL 17  
:|:|:|:|:|:|:  
Db 205 mpsysislgееееvedl 221

RESULT 6  
AAW61347  
ID AAW61347 standard; Protein; 2627 AA.

AC AAW61347;

DT 12-OCT-1998 (first entry)

DE Human telomerase RNA interactive protein-1 (TRIP1).

KW TRIP1; human; telomerase RNA interactive protein-1; cancer; AIDS; ageing; therapy.

XX

OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 1155 /note= "residue 1155 is given as Xaa in Fig 3
FT	(translates as Gln)."
FP	
PN	MO9821343-A1.
PD	
XX	
XX	22-MAY-1998.
PE	
PR	13-NOV-1997; 97WO-US21248.
PR	16-OCT-1997; 97US-0951733.
PR	15-NOV-1996; 96US-0871189.
XX	11-JUN-1997; 97US-0873039.
XX	
	(AMGE-) AMGEN CANADA INC.
	(AMGE-) AMGEN INC.
PI	Harrington LA, Robinson MO;
XX	
DR	WPI; 1998-297946/26.
DR	N-PSDB; AAV27865.
XX	
PT	New nucleic acid encoding human telomerase protein-2 - used for
PT	regulating telomerase activity, e.g. for treating cancer or acquired
PT	immune deficiency syndrome
XX	
PS	
XX	Example 2; Fig 3; 150pp; English.
CC	This polypeptide comprises human telomerase RNA interactive
CC	protein-1 (TRIP1). Its amino acid sequence was deduced from
CC	cDNA clones (see AAV27865) obtained from a colon tumour cell line
CC	LIM1863 cDNA library. The invention relates to novel genes
CC	encoding components of the telomerase enzyme complex, including
CC	TRIP1 and telomerase protein 2 (TP2, see also AAW61349). These
CC	polypeptides may be useful as therapeutic agents in those cases
CC	where increasing TRIP1 activity or TP2 activity is desired, e.g.
CC	for treatment of HIV infection, AIDS and ageing disorders. In
CC	situations in which TRIP1 and/or TP2 activity is to be decreased,
CC	such as in cancer cells in which TRIP1 activity and/or TP2 activity
CC	is elevated, TRIP1 and/or TP2 may serve as a target to identify a
CC	molecule which inhibits activity, or which decreases or inhibits
CC	the protein-protein interaction of TRIP1 and TP2, or the binding of
CC	either TRIP1 or TP2 to telomerase RNA. Alternatively, ex vivo or
CC	in vivo gene therapy may be used to administer TRIP1 or TP2
CC	antisense molecules, or DNA constructs may serve to disrupt or
CC	enhance TRIP1 and/or TP2 expression in cells, and to create
CC	dominant negative inhibitors of TRIP1 or TP2.
XX	
SO	Sequence 2627 AA;
OY	
Db	Query Match 44.2%; Score 42; DB 19; Length 2627; Best Local Similarity 47.1%; Pred. No. 1.8e+02; Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0. 1 VPSFIRAHDESEVDL 17 :      :    205 mpsylslgseeevedl 221
RESULT 7	
ID	AAR9797 standard; Protein: 713 AA.
AC	AAR9797;
DT	26-NOV-1996 (first entry)
XX	
DE	Lysine decarboxylase (W3110) of E.coli.
XX	.

KW	Lysine decarboxylase; L-lysine; cada; transformant; deletion;
KW	mutant.
OS	Escherichia coli.
PN	WO9617930-A1.
PD	13-JUN-1996.
XX	
PF	05-DEC-1995; 95WO-JP02481.
XX	
PR	09-DEC-1994; 94JP-0306386.
PA	(AJIN ) AJINOMOTO CO INC.
PL	Kikuchi Y, Kojima H, Suzuki T;
DR	WPI: 1996-287175/29.
XX	N-PSTDB: AAT34583.
PT	Production of L-lysine by culture of transformant Escherichia - in which expression of new lysine decarboxylase gene and or Cada gene has been partly or wholly suppressed
PS	Claim 1; Page 24-27; 45pp; Japanese.
CC	Transformant forms of Escherichia species (e.g. E.coli), in which the expression of the W3110 lysine decarboxylase gene (AAT34583), and/ or the cada gene (AAT34584), have been partly or wholly suppressed by deletion of part or all of the gene may be used for the efficient production of L-lysine when cultured.
SQ	Sequence 713 AA;
Query Match	43.7%; Score 41.5; DB 17; Length 713;
Best Local Similarity	45.0%; Pred. No. 47;
Matches 9; Conservative 3; Mismatches 5; Indels 3; Gaps 1;	
QY	1 VPSYSFIRAH--DSEVODL 17 :   :       :
Db	77 lplyafinthsundsvqdm 96
RESULT 8	
ID	AAE06785
AC	AAE06785 standard; Protein; 170 AA.
XX	
DT	16-OCT-2001 (first entry)
DE	Human dual-specificity phosphatase (DSP)-12 partial protein.
XX	
KW	Human; dual-specificity phosphatase; DSP-12; dual phosphorylation motif; mitogen-activated protein kinase; MAP-kinase; graft-versus-host disease; GVHD; Duchenne muscular dystrophy; therapy; cancer; autoimmune disease; allergy; metabolic disease; cell growth; cell proliferation; cytosolic; cell cycle abnormality; cell differentiation; antiallergic; muscular; immunosuppressive.
XX	
OS	Homo sapiens.
PN	WO200157221-A2.
XX	
PD	09-AUG-2001.
XX	
PF	01-FEB-2001; 2001WO-US03429.
XX	
PR	02-FEB-2000; 2000US-0179886.
XX	
PA	(CEPT-) CEPTYR INC.
XX	



P1	Luche RM	Wei B;
XX		
DR	WPI: 2001-486887/53.	
XX		
PT	New isolated dual-specificity phosphatase polypeptide for treating	
PT	cancer, graft-versus-host disease, autoimmune diseases, allergies,	
PT	metabolic diseases, abnormal cell growth and abnormal cell	
PT	proliferation -	
XX		
PS	Example 1; Fig 6; 81pp; English.	
XX		
CC	The present sequence is human dual-specificity phosphatase (DSP)-12	
CC	partial protein. Inactivation of mitogen-activated protein kinase (MAP-	
CC	kinase) is mediated by dephosphorylation at a dual phosphorylation motif	
CC	by DSP which is referred to as MAP-kinase phosphatase. An agent that	
CC	modulates DSP is useful for treating a disorder selected from Duchenne	
CC	muscular dystrophy, cancer, graft-versus-host disease (GVHD), autoimmune	
CC	diseases, allergies, metabolic diseases, abnormal cell growth, abnormal	
CC	cell proliferation and cell cycle abnormalities. DSP is useful for	
CC	identifying antibodies and other agents that inhibit DSP-12 and/or	
CC	DSP-13 activity. DSP and the agents identified are useful for	
CC	modulating cell proliferation, differentiation and survival. DSP is	
CC	useful in screening assays for modulators of enzyme activity and	
CC	substrate binding and for dephosphorylating a substrate of DSP-12 or	
CC	DSP-13.	
SQ	Sequence 170 AA;	
XX		
Query Match	43.2%; Score 41; DB 22; Length 170;	
Best Local Similarity	47.1%; Pred. NO. 11;	
Matches	8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;	
OY	3 SYSFIRAHDSFYODLIA 19	
	:     :         :	
Db	72 aymnirvydeetdlia 88	
RESULT 9		
AAR40918		
ID	AAR40918 standard; Protein; 344 AA.	
XX		
AC	AAR40918;	
XX		
DT	21-FEB-1994 (first entry)	
XX		
DE	Bacillus subtilis inositol dehydrogenase.	
XX		
KW	Recombinant protein production; inositol dehydrogenase. Bacillus subtilis.	
PM	JP05192163-A.	
XX		
PD	03-AUG-1993.	
XX		
PE	30-SEP-1991; 91JP-0252073.	
XX		
PR	30-SEP-1991; 91JP-0252073.	
XX		
PA	(SUNR ) SUNTORY LTD.	
XX		
XX	WPI: 1993-277476/35.	
DR	N-PSDB; AAO48539.	
XX		
PT	Inositol dehydrogenase gene - isolated from e.g. bacteria	
PT	belonging to Bacillus subtilis, in large amt.	
XX		
PS	Claim 2; Page 2; 18pp; Japanese.	
XX		
CC	The Bacillus subtilis inositol dehydrogenase gene has been cloned	
CC	and sequenced. Microorganisms transformed by the coding sequence can	
CC	be cultured to produce large amounts of recombinant inositol	

CC	dehydrogenase.
XX	
SQ	Sequence 344 AA;
QY	6 FIRAHDSVDDLI 18   :   :   :   276 fvaaydveldgdl 288
Dd	
RESULT 10	
ID	AAE06772 standard; Protein: 552 AA.
XX	
AC	AAE06772:
XX	
DT	16-OCT-2001 (first entry)
XX	
DE	Human dual-specificity phosphatase (DSP)-12 protein.
XX	
KW	Human; dual-specificity phosphatase; DSP-12; dual phosphorylation motif; mitogen-activated protein kinase; MAP-kinase; graft-versus-host disease; GVHD; Duchenne muscular dystrophy; therapy; cancer; autoimmune disease; allergy; metabolic disease; cell growth; cell proliferation; cytostatic; cell cycle abnormality; cell differentiation; antiallergic; muscular; immunosuppressive.
KW	
OS	Homo sapiens.
XX	
FT	Key Location/Qualifiers
FT	Domain 249..269
FT	/note= "Active site domain"
FT	Misc-difference 386
FT	/note= "Amino acid Thr is present at this location in the sequence shown in sequence listing of the specification"
XX	
PN	WO200157221-A2.
XX	
PD	09-AUG-2001.
XX	
PF	01-FEB-2001; 2001WO-US03429.
XX	
PR	02-FEB-2000; 2000US-0179886.
XX	
PA	(CEPT-) CEPTVR INC.
PI	Luche RM, Wei B;
XX	
DR	WPI: 2001-488887/53.
DR	N-PSDB; AAD12965.
XX	
PT	New isolated dual-specificity phosphatase polypeptide for treating cancer, graft-versus-host disease, autoimmune diseases, allergies, metabolic diseases, abnormal cell growth and abnormal cell proliferation -
PT	
XX	
PS	Claim 1; Fig 2; 81pp; English.
XX	
CC	The present sequence is human dual-specificity phosphatase (DSP)-12 protein. Inactivation of mitogen-activated protein kinase (MAP-kinase) is mediated by dephosphorylation at a dual phosphorylation motif by DSP which is referred to as MAP-kinase phosphatase. An agent that modulates DSP is useful for treating a disorder selected from Duchenne muscular dystrophy, cancer, graft-versus-host disease (GVHD), autoimmune diseases, allergies, metabolic diseases, abnormal cell growth, abnormal cell proliferation and cell cycle abnormalities. DSP is useful for identifying antibodies and other agents that inhibit DSP-12 and/or DSP-13 activity. DSP and the agents identified are useful for
CC	

CC modulating cell proliferation, differentiation and survival. DSP is  
CC useful in screening assays for modulators of enzyme activity and  
CC substrate binding and for dephosphorylating a substrate of DSP-12 or  
CC DSP-13.  
XX  
SQ Sequence 552 AA;

Query Match 43.2%; Score 41; DB 22; Length 552;  
Best Local Similarity 47.1%; Pred. No. 43;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 3 SYSFIRAHDEYVDLIA 19  
: | | | | | | | | | |  
Db 214 aynhlrvydeetdlla 230

## RESULT 11

AAE07043 standard; Protein: 552 AA.

AAE07043;

16-OCT-2001 (first entry)

Human dual-specificity phosphatase (DSP)-12 mutant protein, C253S.

Human; dual-specificity phosphatase; DSP-12; dual phosphorylation motif;  
mitogen-activated protein kinase; MAP-kinase; graft-versus-host disease;  
GVHD; Duchenne muscular dystrophy; therapy; cancer; autoimmune disease;  
allergy; metabolic disease; cell growth; cell proliferation; cytostatic;  
cell cycle abnormality; cell differentiation; antiallergic; muscular;  
immunosuppressive; mutant; mutein; variant.

Homo sapiens.  
Synthetic.

Key Location/Qualifiers

Misc-difference 253 /note= "Wild type Cys substituted with Ser"

WO200157221-A2.

09-AUG-2001.

01-FEB-2001; 2001WO-US03429.

02-FEB-2000; 2000US-0179886.

(CEPT ) CEPTYR INC.

Luche RM, Wei B;

WPI; 2001-488887/53.

New isolated dual-specificity phosphatase polypeptide for treating  
PT cancer, graft-versus-host disease, autoimmune diseases, allergies,  
PT metabolic diseases, abnormal cell growth and abnormal cell  
PT proliferation

Disclosure; Page -: 81pp; English.

The present sequence is human dual-specificity phosphatase (DSP)-12  
mutant protein, C253S. Inactivation of mitogen-activated protein kinase  
(MAP-kinase) is mediated by dephosphorylation at a dual phosphorylation  
motif by DSP which is referred to as MAP-kinase phosphatase. An agent  
that modulates DSP is useful for treating a disorder selected from  
CC Duchenne muscular dystrophy, cancer, graft-versus-host disease (GVHD),  
CC autoimmune diseases, allergies, metabolic diseases, abnormal cell growth,  
CC abnormal cell proliferation and cell cycle abnormalities. DSP is useful  
CC for identifying antibodies and other agents that inhibit DSP-12 and/or  
CC DSP-13 activity. DSP and the agents identified are useful for  
modulating cell proliferation, differentiation and survival. DSP is

CC useful in screening assays for modulators of enzyme activity and  
CC substrate binding and for dephosphorylating a substrate of DSP-12 or  
CC DSP-13.  
CC Note: The present sequence is not shown in the specification, but is  
CC derived from the DSP-12 sequence (AAE06772) given as SEQ ID NO: 2 in  
CC figure 2.  
XX  
SQ Sequence 552 AA;

Query Match 43.2%; Score 41; DB 22; Length 552;  
Best Local Similarity 47.1%; Pred. No. 43;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 3 SYSFIRAHDEYVDLIA 19  
: | | | | | | | | | |  
Db 214 aynhlrvydeetdlla 230

## RESULT 12

AAE07043 standard; Protein: 662 AA.

AAE07043;

26-SEP-2001 (first entry)

C glutamicum protein fragment SEQ ID NO: 6490.

Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
organic acid synthesis.

Corynebacterium glutamicum.

EP1108790-A2.

20-JUN-2001.

18-DEC-2000; 2000EP-0127688.

16-DEC-1999; 99JP-0377484.

07-APR-2000; 2000JP-0159162.

03-AUG-2000; 2000JP-0280988.

(KYOW ) KYOWA HAKKO KOGYO KK.

Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

Tateishi N, Senoh A, Ikeda M, Ozaki A;

WPI; 2001-376931/40.

N-PSDB; AAH67955.

Claim 17; SEQ ID NO: 6490; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein  
sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
are useful for identifying the mutation point of a gene derived from a  
mutant of coryneform bacterium, measuring expression amount and  
analysing the expression profile or expression pattern of a gene derived  
from coryneform bacterium, and identifying a homologue of a gene derived  
from coryneform bacterium. Coryneform bacteria are useful for producing  
amino acids, nucleic acids, vitamins, saccharides and organic acids,  
particularly L-lysine. The present sequence is a protein described  
in the exemplification of the invention.  
Note: The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from the  
European Patent Office.

SQ Sequence 662 AA;

Query Match 43.2%; Score 41; DB 22; Length 662;  
 Best Local Similarity 53.3%; Pred. NO. 53;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 5 SFIRAHSEVODLIA 19  
 | : | | | | | : |  
 Db 415 siekhdspaqdlva 429

RESULT 13  
 ID AAE04833 standard; Protein: 1049 AA.  
 AC AAE04833;  
 DT 10-SEP-2001 (first entry)  
 XX Human SGP006 phosphatase polypeptide.

Human; SGP006 phosphatase polypeptide; phosphatase-related disease;  
 immune-related disorder; ocular disease; organ transplant rejection;  
 infection; diabetes; pain; sexual dysfunction; Alzheimer's disease;  
 metabolic disorder; haematopoietic cancer; mood disorder; cardiac;  
 Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;  
 cardiovascular disease; brain; neuronal-associated disease; dyskinesia;  
 attention disorder; cognitive disorder; psychotic disorder; cytostatic;  
 neurological disorder; vincristine; neotropic; cerebroprotective; therapy;  
 neuroprotective; antibacterial; vulvar; tranquilliser; antiasthmatic;  
 hypotensive; immunosuppressive; antipruritic; analgesic; hypertensive;  
 antifungal; dual specificity phosphatase; DSP; MAP kinase phosphatase;  
 MKP; mgraine; chromosome 12q21.3-q22.

XX Homo sapiens.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 1..173  
 FT /label= Catalytic\_domain  
 FT Domain 308..446  
 FT /label= Phosphatase\_domain

XX MO200146394-A2.  
 XX 28-JUN-2001.  
 XX 21-DEC-2000; 2000MO-US34736.  
 XX 21-DEC-1999; 99US-0173255.  
 XX 28-DEC-1999; 99US-0175766.  
 XX 25-JAN-2000; 2000US-0178078.  
 XX 31-JAN-2000; 2000US-0179301.  
 PR  
 XX  
 PA (SUGEN-) SUGEN INC.  
 XX  
 PI Plowman GD, Martinez R, Whyte D, Manning G, Sudarsanam S, Hill RJ;  
 PI Flanagan P;  
 XX  
 DR WPI: 2001-418058/44.  
 DR N-PSDB: AAD09491.  
 XX  
 PT Novel phosphatase polypeptide useful for treating cancers,  
 PT immune-related diseases and disorders, cardiovascular disease, brain or  
 PT neuronal-associated diseases and metabolic disorders -  
 XX  
 PS Claim 7; Fig 2; 186pp; English.  
 XX  
 CC The present invention relates to phosphatase polypeptides, nucleotide  
 CC sequences encoding them, as well as various products and methods useful  
 CC for the diagnosis and treatment of various phosphatase-related diseases  
 CC and conditions. Substance that modulates the activity of phosphatase  
 CC polypeptide is used to treat immune-related diseases and disorders,

CC cardiovascular disease, brain or neuronal-associated diseases and  
 CC metabolic disorders, including cancers of tissues, cancers of  
 CC haematopoietic origin, diseases of central and peripheral nervous  
 CC system, Alzheimer's disease, Parkinson's disease, multiple sclerosis,  
 CC amyotrophic lateral sclerosis, viral infections, infections caused by  
 CC prions, bacteria and fungi, ocular diseases, diabetes, migraines, pain,  
 CC sexual dysfunction, mood disorders, attention disorders, cognitive  
 CC disorders, hypotension, hypertension, psychotic disorders, neurological  
 CC disorders, dyskinesias and organ transplant rejection. The present  
 CC amino acid sequence is human SGP006 phosphatase polypeptide. This  
 CC sequence is classified as dual specificity phosphatase (DSP) and MAP  
 CC kinase phosphatase (MKP). SGP006 gene maps to chromosomal position  
 CC 12q21.3-q22.  
 XX  
 SQ Sequence 1049 AA;

Query Match 43.2%; Score 41; DB 22; Length 1049;  
 Best Local Similarity 47.1%; Pred. NO. 91;  
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 3 SYSFIRAHSEVODLIA 19  
 : | | | | | : |  
 Db 354 ayhmrvydeetdlia 370

RESULT 14  
 AAY19780  
 ID AAY19780 standard; Protein: 1120 AA.  
 XX  
 AC AAY19780;  
 XX  
 DT 19-JUL-1999 (first entry)  
 XX  
 DE B. burgdorferi antigenic protein, t12.aa.  
 XX  
 KM Antigenic protein; vaccine; Lyme disease; infection; detection.  
 XX  
 OS Borrelia burgdorferi.  
 XX  
 PN WO9859071-A1.  
 XX  
 PD 30-DEC-1998.  
 XX  
 PF 18-JUN-1998; 98WO-US12718.  
 XX  
 PR 03-SEP-1997; 97US-0057483.  
 PR 20-JUN-1997; 97US-0050359.  
 PR 22-JUL-1997; 97US-0053344.  
 PR 22-JUL-1997; 97US-0053377.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (MEDI-) MEDIMUNE INC.  
 XX  
 PI Choi GH, Ewin AL, Hanson MS, Lathigra R;  
 XX  
 DR WPI: 1999-189980/16.  
 DR N-PSDB: AAX61476.  
 XX  
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop  
 PT products for the diagnosis, prevention and treatment of diseases  
 PT caused by Borrelia, particularly Lyme disease  
 XX  
 PS Claim 12; Page 57; 275pp; English.  
 XX  
 CC This sequence represents a Borrelia burgdorferi (Bb) protein of the  
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
 CC can be used in vaccines for eliciting protective antibodies to members of  
 CC the Borrelia genus, particularly for the use against Lyme disease in  
 CC humans and animals. They can be used for preventing or attenuating an  
 CC infection caused by a member of the Borrelia genus. The products can also  
 CC be used for detection of members of the Borrelia genus.

SQ Sequence 1120 AA;

Query Match 43.2%; Score 41; DB 20; Length 1120;

Best Local Similarity 46.7%; Pred. No. 99;  
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 4 YSFIRAHDSVQDLI 18  
|||: |||: |  
Db 740 ysfnkkydsdvkslv 754

RESULT 15

AAV58277  
ID AAV58277 standard; Protein; 1123 AA.

XX AAV58277;

27-MAR-2000 (first entry)

DE Heliothis armigera entomopoxvirus (HaEPV) DNA polymerase.

KM Entomopoxvirus; EPV; insect virus; intergenic region; heterologous DNA;  
KW Insecticide; recombinant protein production; DNA polymerase; HaEPV.

OS Heliothis armigera entomopoxvirus.

PN WO9963062-A1.

XX 09-DEC-1999.

XX 28-MAY-1999; 99WO-AU00416.

XX 29-MAY-1998; 98AU-0003805.

PA (CSTR ) COMMONWEALTH SCI & IND RES ORG.

PI Dali DJ;

DR WPI: 2000-072880/06.

DR N-PSDB; AAV58277.

XX Infectious recombinant virus for use as insecticides or for expressing  
PT therapeutic proteins

PS Examples; Fig 1b; 84pp; English.

This sequence represents Heliothis armigera entomopoxvirus (HaEPV)  
DNA polymerase. The invention relates to infectious, spindle  
body-producing entomopox viruses which have heterologous DNA inserted  
in a 771 bp intergenic region between the fusolin and p68 genes.  
CC Recombinant viruses containing heterologous DNA encoding  
CC an insecticidal protein may be used for the control of insect pests.  
CC Alternatively, viruses containing sequences encoding therapeutically  
CC useful proteins such as interferon, human growth hormone or insulin may  
CC be used for recombinant production of these proteins.

SQ Sequence 1123 AA;

Query Match 43.2%; Score 41; DB 21; Length 1123;

Best Local Similarity 58.3%; Pred. No. 99;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 PYSFIRAHDS 13  
| | | | | | |  
Db 684 pdysyllindke 695

Search completed: March 27, 2002, 13:57:53  
Job time: 521 sec